MIND THE GAPS: MAPPING AND MITIGATING EXCLUSIONARY DATA BIAS IN CRISIS INFORMATICS

A Dissertation
Presented to
The Academic Faculty

by

Rachel Samuels

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in the
School of Civil and Environmental Engineering

Georgia Institute of Technology
AUGUST 2020

MIND THE GAPS: MAPPING AND MITIGATING EXCLUSIONARY DATA BIAS IN CRISIS INFORMATICS

Approved by:

Dr. John E. Taylor, Advisor
School of Civil and Environmental Engineering
Georgia Institute of Technology

Dr. Qi Ryan Wang
School of Civil and Environmental Engineering
Northeastern University

Dr. Emily Grubert
School of Civil and Environmental Engineering
Georgia Institute of Technology

Dr. Michael Elliott
School of City & Regional Planning and Public Policy
Georgia Institute of Technology

Dr. Iris Tien
School of Civil and Environmental Engineering
Georgia Institute of Technology

Date Approved: May 27, 2020

For Paul, without whom I never would have walked this path. Thank you for showing me the good that good research can do.

ACKNOWLEDGEMENTS

This has been a long journey, and I have been exceedingly fortunate to have had so many people help me along the way. Although an exhaustive list would likely be longer than the remainder of this dissertation, I would like to thank a few individuals without whom this would not have been possible.

First, thank you to my advisor and committee chair, Dr. John Taylor. Thank you for taking a chance on a complete stranger on Visiting Students' Day (jinx!), on a new student struggling to adapt to a brand new field, and on each of the wild ideas I have presented to you thus far. I appreciate your continued understanding, patience, and support as I stumbled around in the dark, trying to judge the worth of my ideas by their weight and shape. As I look towards my next steps, I am grateful to you and the lab for continually pushing and challenging me to do and be better.

Thank you to my committee members for their guidance and genuine desire to transform my research. Dr. Elliott, thank you for your perspective on how to use the histories and scars of cities to inform my analyses. Dr. Grubert, thank you for your kindness, your insight, and your honesty in improving the broader implications and applications of my work. Dr. Tien, thank you for your unparalleled teachings on data bias and the methods we can use to interrogate it. Dr. Wang, thank you for laying the foundations in the Network Dynamics Lab for studying disasters and for continuing to support and shape the disaster research that followed yours.

Within the lab, I am overwhelmingly grateful for the camaraderie and love that my labmates have given me over the years. Thank you so much to Abby, Susie, Lei, Neda, Praga, Dimitri, and Yuli. Without your constant willingness to review my work, give me honest feedback, and brainstorm with me on solving a seemingly endless stream of problems, I could never have come this far.

I owe so much of any success that I have had to my family. Thank you to my father for his unending faith in me, his wry good humor, and his conviction that I would eventually be where I am now: submitting my PhD thesis. Thank you to my mother for constantly pushing me to do more, for her constant support, and for always encouraging me to go that last extra mile. Thank you to my stepmother for her welcoming heart and for always being willing to listen when I felt like I had run out of options.

My incredible friends are as dear to me as family. They have gone above and beyond the call of duty in encouraging me, supporting me, and forgiving me when I had to cancel a plan or forgot to respond to yet another message. I am so sorry not to be able to call you all by name, so here are but a few who carried me when I would have surely fallen: Kirstie, Emmy, Laura, Jolyn, Chris, Wen, Stephanie, Jolyn, India, Lamb, and Julie. Thank you to the cosplay crew, the Chattahoochee crew, the Washington & Lee crew, the Barbell Club crew, and the EnvE crew. Thank you, thank you, thank you.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS AND SYMBOLS

ABBREVIATIONS

API	Application Program Interface
ASCE	American Society of Civil Engineers
CDC	Centers for Disease Control
CDF	Cumulative Density Function
CV	Cross-Validation
EMDAT	Emergency Events Database
FEMA	Federal Emergency Management Agency
GIS	Geographic Information Science
IIS	Information and Intelligent Systems
K-S test	Kolmogorov-Smirnov test
MAUP	Modifiable Area Unit Problem
NLCD	National Land Cover Database
NOAA	National Oceanic and Atmospheric Administration
NSF	National Science Foundation
PC1	First Principal Component
PCA	Principal Component Analysis
PCR	Principal Component Regression
RMSEP	Root Mean Square Error of Prediction
SVI	Social Vulnerability index
US	United States
VGI	Volunteered Geographic Information
ZCTA	Zip Code Tabulation Area

SYMBOLS

a_p	Number of Tweets (activity) observed in a defined period in a hexagon within the perturbed state
α	Scaling constant for a power law function
β	Power coefficient for a power law function
β_0	Regression intercept term
β_i	Regression coefficient
δ_a	Twitter activity deviation
ϵ	Regression error term
$HighInt_t$	Number of cells described as “High Intensity (Developed)”
km	Kilometers
$LowInt_t$	Number of cells described as “Low Intensity (Developed)” within the tract
$MedInt_t$	Number of cells described as “Medium Intensity (Developed)”
m	Meters
μ_s	Average number of Tweets observed in that area during the steady state

$OpenArea_t$	Number of NLCD cells described as “Open Space (Developed)” within the census tract
PC_1	The first principal component
Pop_{cell_i}	Population represented by a single 30mx30m NLCD raster cell point of NLCD type i within a specific census tract
Pop_t	Population for a given census tract t
$Pop_{total_{tract}}$	Total population within the census tract in which the point is located
ρ	Spearman rank correlation coefficient
σ_s	Standard deviation of the Tweet counts observed during the steady state.
τ	Kendall rank correlation coefficient
$Tract_{count_i}$	Total number of points of type i within the specific census tract
$TwAct_{state}$	Daily Tweets per person in the temporal state being analyzed
WA_i	Weighted average for the specific land type

SUMMARY

Increasingly large numbers of people are living in areas susceptible to catastrophic disasters because of urban sprawl (Allen 2006) and worsening extreme weather patterns from climate change (Adachi et al. 2017). While more severe weather events is becoming more of a certainty than a possibility (Hauer et al. 2016), the extent of the impact on humans and society can be mitigated through improved resource planning and resource agility by increasing real-time information on human location, activity, and responsiveness (Roshan et al. 2016). New and varied forms of information from humans-as-sensors are being utilized in crisis response, and there has been a substantial push towards finding ways of applying data such as social media to emergency responder needs on the ground and at higher levels of decision-making (Lachlan et al. 2016; Reuter et al. 2018; Spence et al. 2016). However, although many potential uses for social media information have been identified across crisis informatics, a number of both ethical and computational biases have been identified as well. An important area for research is identifying these biases, the effects they have on disadvantaged populations, and how to mitigate that bias in the growing body of work designed to utilize social media in crisis situations. Within this dissertation, I describe three studies that identify, define, and utilize select limitations in social media for crisis response. In the first of these studies, I examine the prevalence and significance of decreases in social media activity from normal state to crisis conditions. Through correlating changes in social media usage and infrastructural damage, I show the importance of considering social media usage drop-offs in crisis identification. In the second, I examine the influence of geographic scale on the statistical reliability of social media data and the correlation between social media and infrastructural damage. By varying the geographic scales at which I aggregate behavior, I show the high sensitivity of social media usage analytics to scale and the consequences of neglecting to incorporate scale into existing research conclusions. For the third, I examine the effect of social vulnerability factors on the presence or absence of social media data during a disaster. By comparing the contribution of social vulnerability factors to social media data availability during a normal state and a crisis state, I show that social vulnerability contributes heavily to a decrease in data in a crisis state that is not present during a normal state. The results of this work inform the reliable extent of social media data and its sensitivity to external factors (i.e. infrastructure damage and the presence of vulnerable populations) and analytical factors (i.e. spatiotemporal scale, aggregation, and bursting behavior). This work is ultimately driven by the need for our cities to improve as disasters are worsening. Social media analytics offer one method of improving our crisis response; however, any new technology holds the danger of leaving certain populations—especially vulnerable populations—behind. By pinpointing disparities in the representational capacities of the data and proposing alternative methods of use, I hope to improve the usability and equity of social media data for crisis response.

CHAPTER 1. INTRODUCTION

1.1 Increasing global risks and at-risk populations

The global cost of disasters and the number of people affected by them have been increasing steadily over the last few decades (EMDAT 2020). Part of this is due to urban sprawl; more people than ever are living in areas, such as the coastline, that are susceptible to weather event-induced disasters (Allen 2006). The urban population in the United States (US) continues to grow, and coastal cities are growing even faster than their inland counterparts. This phenomenon is what has been referred to as the ‘expanding bullseye’ effect: the expanding coastal populations are becoming a larger target for extreme storms to hit (Ashley et al. 2014). Secondarily, an increase in extreme weather patterns is anticipated to be a major consequence of worsening climate change (Adachi et al. 2017). Alongside an increase in weather extremity, hurricanes especially are predicted to behave in more erratic patterns. This includes a decrease in lateral movement, causing the storms to deposit more rain on impacted cities and also to impact cities that have historically been in less danger from powerful storms (Noy 2016). Direct losses from hurricane and flood damages have tripled across the past fifty years, and the Southeast in particular has seen economic impacts outpacing population growth (Gall et al. 2011). As both populations at risk and the severity of those risks continue to increase, finding innovative ways to protect large coastal communities will become more necessary.

One of the critical pieces of this puzzle is improved, agile emergency management. Disasters are notorious for being situations in which reliable, specific information is scarce, and immediate, deadly crises can happen at relatively small scales (Quarantelli 2003;

Wurman and Kosiba 2018). Emergency managers need specific information about the "what, when, where, and who" of these local crises (Yang et al. 2013), and one potential source of that information has been identified as the people living in the impacted community. With the surge in online communication and number of ways people can contribute to news sources and information repositories, the utilization of this non-traditional form of information is becoming both more possible and more necessary.

1.2 Volunteered geographic information

Community- and place-based resilience has long been at the forefront of improving crisis response (Cutter et al. 2008). Finding new ways to utilize the members of a community that are inclined towards pro-social behavior—which has been shown to predominate in disaster situations—is one method of increasing community resilience. As one of the greatest strengths of community members is their knowledge of the impacted place and the people residing there, volunteered geographic information (VGI) has significantly helped emergency management in both the "Response" and "Recovery" stages of disasters (Reuter and Kaufhold 2017). VGI is information contributed by the community regarding the location or condition of local problems or resources. It can include which roads are open, where the power is out, where trees are blocking roads, and calls for help. For instance, one of the most utilized forms of VGI is OpenStreetMap, a site in which people can review satellite footage or offer on-the-ground insight as to local dangers and intact streets (Eckle and de Albuquerque 2015). It was instrumental in the response to the 2010 Haitian earthquake, when information about local needs and available resources was scarce (Zook et al. 2010). In part due to the success of VGI applications such as OpenStreetMap and Ushahidi in 2010, response organizations began looking for ways to incorporate that and other forms of VGI into their official response plans.

Currently, VGI (especially in the form of social media) is being utilized in an official capacity by the American Red Cross, the United Nations in various resilience efforts, the Humanitarian Tracker, and many other nonprofits (Gao et al. 2011; Imran et al. 2014; Lovejoy et al. 2012). In a less official capacity, local emergency managers have taken to monitoring social media streams manually, searching specific hashtags and opening communication pathways through government accounts. These managers also predict an increase in their utilization of VGI through platforms such as social media to both send messages and receive them to improve situational awareness (Hiltz and Kushma 2014). Social media as a form of VGI is notable for containing information from eye-witness accounts and being available in near-real time. Twitter in particular has seen substantial usage due to its ease of access via the Twitter streaming Application Program Interface (API) and the sheer quantity of posts available through that API. The number of Twitter users has been steadily increasing since Twitter was founded in 2006, with 330 million monthly users and almost 150 million daily users (Twitter 2018), so the information contained within its data stream covers a wide range of users and topics. Historically, users were able to choose whether or not to include a geotag—a latitude and longitude pair—within their posts. Currently, users are able to do the same thing with pictures taken through the Twitter app. The prevalence of geotagged information with high location accuracy, the potential of a picture of the situation, and a user that can be contacted for additional information has motivated substantial usage of Twitter data.

Unfortunately, although the quantity of posts has a positive impact on the amount of information Twitter can provide, there is also a lot of unrelated or unreliable data included in the datastream (Starbird et al. 2014). This, in turn, has motivated substantial research projects devoted automatically parsing the Twitter datastream and finding ways to reliably and accurately utilize Twitter posts to aid in crisis mitigation efforts.

The incorporation of VGI has been useful for tracking individuals' mobility and the influence of a disaster on that mobility (Wang et al., 2017), the change of individuals' sentiment in response to different disaster impact levels (Wang and Taylor, 2018), and to identify infrastructure service disruptions using social media data mid-disaster (Fan and Mostafavi 2019). Research using spatiotemporal aggregation to compare two spatial datasets has shown that bursts of social media behavior and disaster-related posts can indicate areas of relatively higher hurricane damage (Kryvasheyeu et al. 2016) and the location of flooding (de Albuquerque et al. 2015). Looking forward, VGI has been proposed for use in digital twin city frameworks for assessing infrastructural vulnerabilities, thus improving disaster resilience and preparedness (Xu et al. 2016), and for improving the situational awareness of emergency responders using the digital twin through integrated text, image, and geopositioning analysis (Fan et al. 2020).

However, across fifteen years of crisis research, the field has focused primarily on where and how social media data can be employed at broad scales (Reuter and Kaufhold 2018). There is considerable need to further investigate the question of where and how social media data can be used *equitably*.

1.3 Equity and actionability in social media usage

Big data research—and social media is a form of big data with its high volume, variety, velocity, and questionable veracity—has often been critiqued for overlooking human variability and for mistaking big data for complete data (Blumenstock 2018; Gandomi and Haider 2015). These two fallacies can also be found intertwined in some aspects of existing crisis informatics, as one of the critical dilemmas with humans-as-sensors analyses is that humans are not reliable sensors. Humans do not transmit consistent, coordinated, or comparable information through public data channels that can be continuously accessed by connected emergency responders or data analysts. The rush to utilize information produced by humans in disasters has neglected to incorporate the diversity of human response and capabilities, impairing proper management and stewardship of that information.

There is a substantial need in research to investigate how the bias inherent in social media and existing methods of social media analyses affect the types of information produced for emergency management. From sample bias to survivorship bias, the biases inherent in who produces what during a crisis need to be explored prior to the development and utilization of social media-parsing applications for crisis informatics. To enable this, the boundaries of use and the consequences of failing to account for these biases need to be defined in the context of emergency identification and response.

1.4 Dissertation framing and structure

Social media is a new and promising form of crisis information that, as a non-traditional form of communication, has seen a surprisingly exponential increase in global use. It is one potential method of improving community resilience in the face of increasing weather-related disasters caused by climate change. New methods of parsing usable information from social media have shown substantial potential for enabling social media's use in crisis response; however, further assessments of where and how social media can be used equitably are necessary. This dissertation aims to improve our understanding of the limits of social media analyses in the context of both our existing methods of analyses and the socioeconomic disparities of social media's use. Below, I detail three studies in which I assess potential biases inherent in the existing methods of examining social media for emergency response.

In Chapter 2, I examine potential survivorship bias in the existing use of activity bursts of social media as a form of crisis identification. Research has shown that bursts of social media activity are positively correlated with infrastructural damage. However, criticisms of that data have shown that some areas with large amounts of damage were silent on social media during the disaster. The significance of that social media silence in crisis informatics analyses and the potential benefits of using it to rapidly assess damage had not been investigated. To address this need, I compared social media activity before Hurricanes Harvey, Irma, and Nate in nine affected cities to social media activity during each hurricane to determine the social media activity deviation. I determined the correlation between infrastructure damage and 1) the absolute activity deviation (the magnitude of the change), and 2) the raw activity deviation (the direction and magnitude of the change). The article presented in Chapter 2 has been published in *Natural Hazards*.

In Chapter 3, I examine the impacts of the modifiable areal unit problem on social media data availability. In identifying crises through social media, contextualizing the data is critical for understanding the magnitude and severity of the crisis. As shown in Chapter 2, the temporal context (i.e. historical data) is important; this chapter analyzes the geographical context. Many detection algorithms identify prominent clusters of activity across varying geographic ranges, and historical contextualization most often occurs at a specific geographic range. To address potential bias in data clustering and aggregation, I investigated the impact of spatial scale on the results of a) activity burst identification, b) historical contextualization, and c) the strength of the correlation between social media activity and infrastructural damage. To do so, I aggregated social media data across multiple geographic scales ranging from 0.25 km^2 to 80 km^2 and analyzed the changes in distributions related to extreme activity changes, the statistical robustness of historical data availability, and the correlation between social media activity and infrastructural damage. The article presented in Chapter 3 has been developed into a journal paper and will be submitted to a journal that specializes in Geographic Information Science (GIS) and its applications to engineering problems.

In Chapter 4, I further explore the concepts shown in Chapter 2 that social media silence during a disaster is more likely to be indicative of crisis than not. I theorized that social vulnerability attributes were likely associated with that silence. If this were the case, usage of social media data without social contextualization would prioritize resource distribution to the least vulnerable instead of the most. I additionally wanted to test whether a decrease in social media presence in areas with more vulnerable populations could be predicted from the pre-crisis period. To examine this, I used the Centers for Disease Control's (CDC's) Social Vulnerability Index (SVI) to determine the percentage of people ascribing to 13 different vulnerability factors in each area, then determined the average social media activity level in a steady state and the social media activity level in those areas during a hurricane. I used principle component analysis to transform the vulnerability factors into uncorrelated variables because of the high degree of multicollinearity in the factors. I then regressed these components to analyze the contribution of the components to the social media activity in the steady state and the perturbed state and compared the results. The article presented in Chapter 4 has been developed into a journal paper and has been published by the American Society of Civil Engineers (ASCE) *Journal of Management in Engineering*.

Finally, Chapter 5 includes a discussion of the overarching contributions of this work to how the crisis informatics field develops and applies social media to emergency management. This work has incorporated insights from the fields of critical GIS, social vulnerability science, emergency management, and information science, and I discuss potential new research that is suggested or could become possible following the conclusions listed in each study.

CHAPTER 2. SILENCE OF THE TWEETS: INCORPORATING SOCIAL MEDIA ACTIVITY DROP-OFFS INTO CRISIS DETECTION[1]

2.1 Abstract

Although extreme events are inevitable, the associated cost to infrastructure and human life is not. We can mitigate these costs through improving the information available to emergency responders during and after crisis events via social media. Recent research has identified a correlation between spikes of Twitter activity and the infrastructural damage incurred during natural disasters. This research, however, overlooks emergencies occurring in areas in which people have lost power, lack the ability to connect to the internet, or, due to differences in social media perceptions, are uncompelled to Tweet during a disaster. To assess the prevalence of Twitter activity decreases and the relative importance of those decreases in detecting areas in crisis, we study crisis-driver Twitter activity deviations from "normal" in nine cities affected by the 2017 Atlantic Hurricane Season. In analyzing more than 1.1 million Tweets across the season, we find that there is a stronger, more significant correlation between infrastructure damage and a metric that prioritizes both increases and decreases in Twitter activity than one that prioritizes only Twitter activity increases. These findings indicate that social media drop-offs could be

[1] This chapter is published in the journal *Natural Hazards* with Neda Mohammadi and John E. Taylor as the co-authors. The citation for the journal article is as follows: Samuels, R., Mohammadi, N., and Taylor, J.E. (2020). "Silence of the Tweets: Incorporating social media activity drop-offs into crisis detection." *Natural Hazards*. https://doi.org/10.1007/s11069-020-04044-2

representative of significant distress, and accounting for the apparent survivorship bias in social media will be critical to the equitable use of social media in crisis applications.

2.2 Background

The 2017 hurricane season broke many records in the United States; it produced more Category 5 hurricanes than any other, broke the record for the strongest hurricane recorded in the Atlantic, and was the costliest hurricane season to date. The ensuing 2018 Atlantic Hurricane Season continued to wreak above-average devastation, becoming the third consecutive season to produce a Category 5 hurricane (National Oceanic and Atmospheric Administration 2018). Unfortunately, the rise of both sea level and global temperatures is expected to continue to contribute to an increase in intensity and frequency of extreme weather events (Noy 2016). Contemporaneously, the number of people living in coastal cities is rapidly increasing, even more so than those cities' landlocked counterparts (Hauer et al. 2016). The increased populations and infrastructure on the coastline have resulted in an 'expanding bullseye' for these increasingly deadly storms to hit (Ashley et al. 2014). As both of these phenomena are predicted to continue and to increase in intensity, emergency professionals and urban planners will need to improve their response capabilities. One aspect of this is to improve the quantity and scope of the information utilized in crisis management.

Actionable information is one of the most necessary yet most difficult to obtain resources in the midst of an ongoing disaster. Especially for events that cover a broad geographic and demographic range, crisis managers rarely have access to timely and specific information (Reuter et al. 2018). There is a broad area of research in the field of crisis informatics that seeks to translate diverse sources of data into actionable information for decision-makers. One of the most popular big data sources that has come to prominence in the last decade is that of social media. Social media has been recognized as a potential source of human network information for fifteen years and researched as a potential source of crisis information for at least ten (Reuter and Kaufhold 2018). Applications monitoring social media data have been used by organizations such as the Red Cross and the United Nations Office for the Coordination of Humanitarian Affairs (Imran et al. 2014). Through utilizing this information as data from a "new" type of sensor—or, rather humans-as-sensors—researchers have been able to expand the quantity and types of crisis information specific to both individual residents and broad geographic regions during disasters.

The types of information produced through social media can include a unique user identifier, the time at which the information was published, the content itself (which can include text and images), the geographic region in which the user posted the information, a link to another site, and information about how the information has spread. Not all posts include all of this information, and some studies use information gleaned through determining connections between different posts or users (Caragea et al. 2014). For the sake of clarity and brevity, we focus on the social media information available through Twitter. Twitter data has been the most commonly utilized humans-as-sensors data in the social media sphere due to the relative ease with which it can be accessed (Steiger et al. 2015), and, as our paper also utilizes that data, that is what we will primarily describe. Additionally, its ease of use does not imply a decreased utility; Twitter has a reported 326 million average monthly users and more than 500 million Tweets per day (Twitter 2018).

The utilization of this information in the field of crisis response has taken many forms. These forms tend to have been built primarily along one of two distinct but intertwining pathways: first, filtering individual posts for directly usable information, and second, analyzing variances in information production for spatial, topical, or temporal closeness (Wang et al. 2016; Weiler et al. 2016; Resch et al. 2018). Research utilizing the first seek explanations of need, resource availability, blocked roads, and the emergence of new disasters (Cameron et al. 2012; Purohit et al. 2014). The second utilizes broad-scale aggregated posting behavior changes like sudden increases in a specific topic or sudden bursts of activity in a localized area to inform the credibility of a disaster's occurrence or the extent of a disaster's influence, such as the extent of flooding near the River Elbe (Herfort et al. 2014; de Albuquerque et al. 2015). By interpreting both the individual Tweets and the broader Twitter stream, the spatial, temporal, and topical characteristics of the data have been employed to develop models that can detect emergencies (Imran et al. 2015), identify emergency types (Xu et al. 2016), rapidly assess areas experiencing hurricane-related damage or severe conditions (Guan and Chen 2014), detect resource availability and need (Purohit et al. 2014), identify human sentiment in real time, correlate human sentiment with disaster intensity (Wang and Taylor 2018), detect the emergence of new, unspecified disasters (Wang and Taylor 2019), and characterize resilience and community recovery through mobility patterns (Spence et al. 2015).

That said, although great strides have been made since the conception of crisis informatics and the availability of more big data, we are still stretching and testing the limits of what this data can do and where it is applicable. Big data as a whole has been routinely criticized for increasing the existing disparities between privileged and underprivileged groups, and social media's user demographics are far from broadly representative (Blumenstock 2018). Especially in the world of crisis informatics, understanding the scope of the diversity of human interactions with social media in crises will be critical to mitigating these harmful data biases. Within this paper, we seek to explore the survivorship aspect of the data bias introduced through the analysis of social media data during a crisis.

First, we note that data bias is introduced when a system's characteristics are described by a reduced set of information. This reduction in information happens when analyzing individual posts, as their context in the broader stream is lost, and when analyzing the broader stream as an aggregate of individual posts, as the text-specific information is lost. Most recent applications use a combination of the two methods in tandem to increase data richness, and many studies have been performed analyzing data streams across different phases of a disaster, like preparedness, response, recovery (Yang et al. 2013; Bakkensen et al. 2017; Zou et al. 2018a). Additionally, especially when using humans-as-sensors, there is plenty of bias in which humans choose to contribute to the data stream at any given time. The demographics of Twitter are fairly well understood, and the demographics of vulnerable populations contributing to the data stream have additionally been explored (Zou et al. 2018b; Wang et al. 2019). However, as far as the authors are aware, no social media-based crisis response applications have considered potential data loss created by the transition of the sensors—in this case, the humans—from a non-crisis state to a crisis state. Without incorporating the possible effects of state-transition data loss, the data that exists during a crisis is prioritized, and those who have dropped out of the stream are neglected.

Human sensors are distinct from physical sensors, like flood height monitors, because we are rarely aware of when or how the sensors are "turned off". Especially in the context of social media crisis research, which has generally operated on the idea that people Tweet more in response to a non-normal stimulus due to the identification of Twitter bursts (Kryvasheyeu et al. 2016; Fan and Mostafavi 2019), the identification of people who have stopped Tweeting is a sorely lacking area of investigation. The field has continually increased the contextualization of the used data, but when we only work with visible data, we could be missing the neglected populations that are going dark instead. Shelton et al. identify this darkness as a potential problem during Hurricane Sandy when they found that 50% of deaths in New York occurred in an area with relative social media silence (Shelton et al. 2014). Similarly, Xiao et al. highlighted the potential influence of hurricanes on the "digital divide" with respect to lower socioeconomic groups having more limited access to technologies (Xiao et al. 2015). These findings call for the need to characterize the extent of these geographic "data shadows" and to define their impact on the capacity of social media to identify localized areas of hurricane damage. There is also a research gap with respect to how these shadows change from non-perturbed to perturbed state.

Similarly, although we know that these data shadows exist and shroud some local crises, research has not shown how to see through them. Twitter data is unlikely to ever give us actionable information from people who do not have a Twitter account, but we may not be limited to those people who are only Tweeting during the storm. People that have lost the compulsion or ability to Tweet may, in fact, be in more danger than those that are still Tweeting. To test this, we need to understand if these drop-offs (and so potentially these people) can be identified through different methods of analysis beyond in situ Twitter activity or topic bursts. Thus, we first devise an approach to characterize and identify the changes in Twitter activity from a baseline state. Then, using that approach, we test the following hypothesis:

H1: A crisis metric that prioritizes both social media activity increases and drop-offs is more strongly and significantly correlated with infrastructure damage than one that only prioritizes activity increases.

In testing this hypothesis, we determine whether prioritizing social media activity drop-offs could aid in the identification of extreme hurricane damage. If they are, we need to include or acknowledge them in our crisis response applications so that silenced populations are not disproportionately ignored. In this paper, we report correlations between infrastructural damage and the magnitude of social media activity change from a defined baseline, or steady state. By focusing on social media posting activity changes from a steady state, and thus incorporating social media signal drop-offs, we hope to demonstrate how to increase the usability of social media data for emergency responders while reducing the risk of ignoring areas that do not have the ability or compulsion to Tweet.

2.3 Methods

We analyzed the Twitter posting behavioral shifts from non-crisis to crisis conditions and the incurred infrastructural damage for ten regions across the southeastern United States during four named storms in the 2017 Atlantic Hurricane Season. These regions include two affected by Hurricane Harvey (Beaumont, Texas and Houston, Texas), five affected by Hurricane Irma (Tampa, Florida; Miami, Florida; the Florida Keys; Fort Myers, Florida; and Jacksonville, Florida), one affected by Hurricane Maria (the U.S.

territory of Puerto Rico), and two affected by Hurricane Nate (Biloxi, Mississippi and Mobile, Alabama). The counties associated with those regions (determined through defining the Metropolitan Statistical Areas of each city) and the damage distribution of the four hurricanes are shown in **Figure 1** and **Figure 2** respectively. The damage distribution is represented through Federal Emergency Management Agency (FEMA) Building Level Damage Assessments performed two to five days following the hurricanes' landfalls.

The residential populations of the studied regions range from 70,000 to almost three million, and the number of Tweets produced per day in each region ranged from 200 to 500 in the Florida Keys to 10,000 to 15,000 in Houston, Texas. From this, we calculated a broad spectrum of Twitter penetration (number of Tweets per person) and Twitter usage seasonality (how Twitter activity varies from day to day) within the regions. The lowest Twitter penetration in the steady state was observed as 0.0017 in Biloxi, MS, and the highest penetration during the steady state was 0.0081 in Houston, TX. During the perturbed state, the lowest was 0.0001 in the Florida Keys, and the highest was 0.065 in Miami, Florida. In terms of variation in activity, we observed increases in average Twitter usage on the order of 40-70% on weekends for each region. As our total data set included a number of US holidays such as July 4th and Labor Day, we also noted that social events increased the number of produced Tweets. The hurricanes that impacted these cities, while all dangerous storms, also differed in wind speed, rainfall, amount and speed of flooding, and, ultimately, the broad scale infrastructure damaged caused. Crisis management differed as well, as some cities had evacuation orders well in advance—more than 6 million people evacuated for Hurricane Irma (Reynard and Shirgaokar 2019)—while others did not have widespread evacuation orders at all (Milliner et al. 2018).

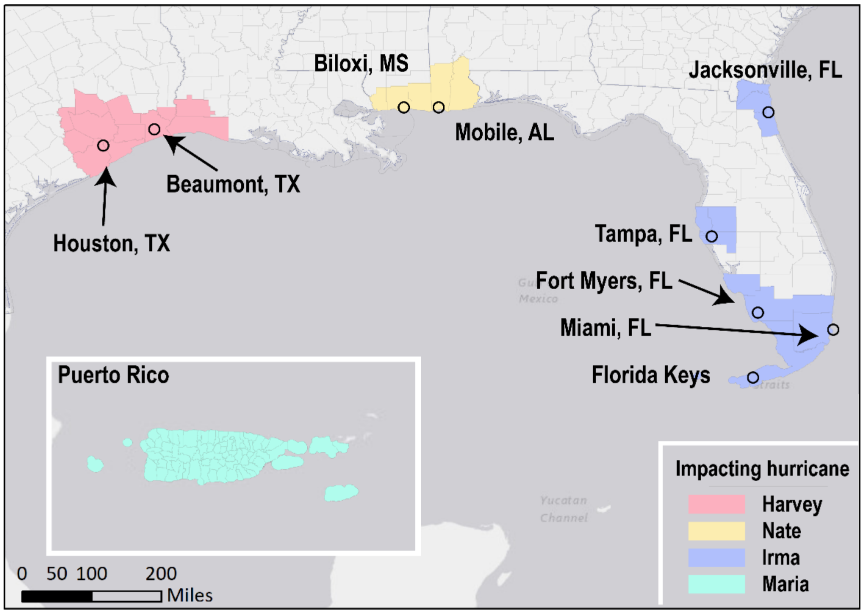

Figure 1. The geographic extent of the study area. The cities of interest are indicated with black circles. Areas are colored according to which hurricane's impact was studied.

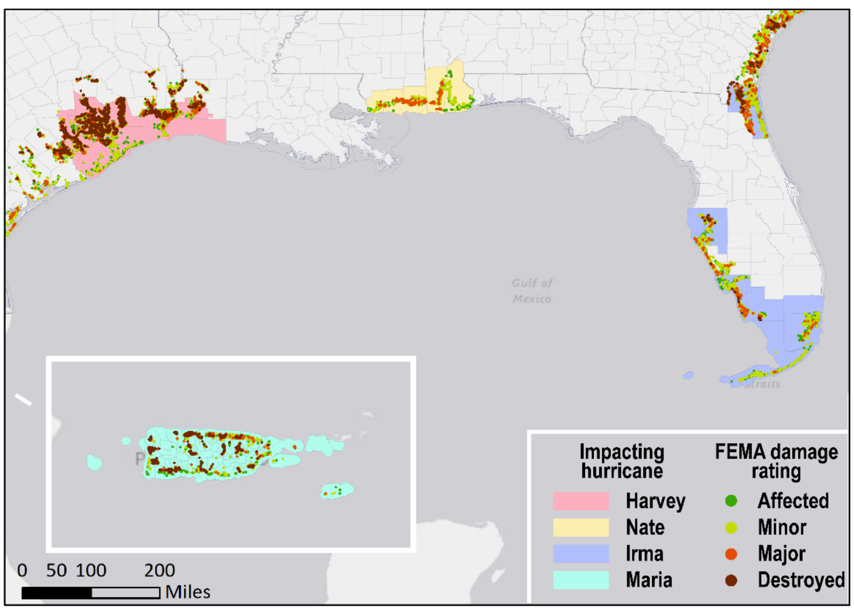

Figure 2. The extent of the FEMA Building Level Damage Assessments across the southeastern US for the 2017 Hurricane Season. Each dot indicates a building that has been assessed. The damage scale transitions from dark green (affected) to red (destroyed).

With our areas of interest defined, we proceeded to acquire the Twitter posts from those regions and filter them into the temporal and spatial segments defined below.

2.3.1 Twitter data acquisition

We streamed Twitter data through the public Twitter Application Programming Interface (API) (https://developer.twitter.com) for the duration of the study. The API timestamps each Tweet as it is posted, and the Tweet and its attributes (e.g., text, user identifier, coordinates of the Tweet, etc.) are pushed in real time to a server in the lab, where the data is stored permanently. Following our Twitter collection during the 2017 Atlantic Hurricane Season, we filtered those Tweets first for those geolocated in the southeastern United States and the Gulf of Mexico, and secondarily for the Metropolitan Statistical Areas for each of the regions outlined above. Each instance of Tweet data includes the location of the Tweet in decimal degrees, the author, the time of posting, and the Tweet's text content. For this analysis, we did not consider retweets, and we did not consider Tweets geolocated through the location-attribute designated in the originator's profile. We additionally processed the Tweets' texts to filter out bots using keywords we built through testing with OSoMe's Botometer tool (Indiana University 2018) and spot-checked the total dataset for location accuracy.

It should be noted here that, with respect to Puerto Rico and Hurricane Maria, we unfortunately found an extremely and unusually high percentage of Tweets recorded post-landfall that appeared to be falsely geolocated on the island. Due to the reported near-total decimation of the power and communications structure on the island likely reducing true posts and the extreme prevalence of incorrectly geotagged data associated with Puerto Rico post-landfall (Pasch et al. 2019), the data for Hurricane Maria was ultimately not considered in the final analysis. The implications of this are explored further in the Discussion section.

Following the text-based filtration, we processed approximately 1.1 million Tweets into the following temporal and spatial segments.

2.3.2 Temporal segmentation

Having identified our process for collecting relevantly located and produced data, we needed to define the time periods from which to draw data. Our stated goals required us to specifically analyze changes from steady state Twitter posting behavior to perturbed state Twitter posting behaviors. The perturbed state period was easiest to define: we considered the perturbed state to consist of the day before landfall and the week following landfall in order to encompass most of the 'response' and 'recovery' phases of the hurricane (Yang et al. 2013). In defining the beginning of the steady state, we wanted to avoid potential activity fluctuations that occurred in anticipation of hurricane damage in the preparedness phase. To do this, we allowed for a transitionary state that we defined as the day the hurricane was identified in the Atlantic until the day before landfall. As for the steady state's duration, we followed the recommendations of Toepke, who identified that defining a reliable baseline for Twitter activity requires four to six weeks of data (Toepke 2018a). A greater time period could be influenced by the seasonality of Twitter usage, and a smaller time period would be more susceptible to influence from outliers. The flow of Twitter data has been shown to be steadily increasing from its founding through 2017, although its growth rate has since leveled off (Twitter 2018). Thus, to minimize the influence of outliers while still generating a robust set of data, our steady state for each region thus consisted of the six-week period prior to when the affecting hurricane was identified in the Atlantic.

With our broad-scale temporal states defined, we needed to assess the temporal scale at which we would perform our analysis within those states. Twitter data varies substantially throughout the day; there are substantially fewer Tweets at 3 a.m. than at 5 p.m., and we needed to define steady states that could account for that quotidian variation while still amassing enough data to develop a consistent baseline. Although we accumulated data in hourly packages, the Twitter activity deviation for individual hours across the steady state was too high. After assessing the rate changes in Twitter activity during the steady state and accounting for changes across time zones, we ultimately defined six temporal categories into which we sorted our data. The data was classified as either "weekday" or "weekend", then sorted into one of the following time windows: 12:00 a.m. to 7:59 a.m., 8:00 a.m. to 3:59 p.m., and 4:00 p.m. to 11:59 p.m. The six temporal categories are shown in **Table 1**. Data acquired during the steady state that was sorted into each of these six temporal categories were utilized to develop six steady state distributions, and the acquired perturbed state data were then compared to their respective steady state distribution.

Table 1. Description of the temporal segmentation for the Twitter data.

Day of the week	12am – 8am	8am – 4pm	4pm – 12pm
Weekday (Mon – Fri)	Category 1	Category 2	Category 3
Weekend (Sat – Sun)	Category 4	Category 5	Category 6

2.3.3 Spatial segmentation

Emergency responders have indicated a need for emergency location, severity, and magnitude information across the spatiotemporal breadth of the crisis (Mason et al. 2017). It is therefore important to be able to identify areas of great need on a fine resolution scale and to understand that need relative to the surrounding areas. However, with any aggregation of individual data points, the aggregation reduces the granularity and spatial accuracy of the data and increases the uncertainty and variability of the aggregation's ability to represent the points it encompasses (Bian and Butler 1999). We had two determinations to make for our methods of aggregation: the size and the shape of our areas. Although census tract boundaries have accrued a long and storied history within any research field involving people, they have also accrued a substantial amount of criticism (Grubesic and Matisziw 2006). Census tracts vary widely in size and shape, ranging from less than 0.1 square kilometers (km^2) to more than 800 km^2, and are defined using socially-influenced boundaries (Fotheringham and Wong 1991; Portnov et al. 2007). Inner cities have much smaller census tracts than broad rural areas, and the conflux of the modifiable areal unit problem and the geographic relationships of social vulnerability factors could heavily influence the results of an analysis with a prominent social aspect. Additionally, we should note that we were able to stretch beyond the limits of census tracts because we were not normalizing the collected Twitter activity by population but rather by historic Twitter activity.

As such, we decided to generate uniform hexagons in which to spatially segment our data. Hexagons have been noted to preserve inherent spatial variations and neighboring relationships and to improve the scalability of the grid (Carr et al. 1992; Polisciuc et al.

2016; Shelton et al. 2014). In terms of the size of the hexagons, the greatest analytical concern was the potential tradeoff between location specificity and sufficient quantities of data. Potter et al. addressed this issue in the field of landscape ecology by comparing data resolution for different types of data at different scales (Potter et al. 2016). We slightly modified their methods of analyzing the how different hexagon sizes resulted in different distributions of event severity (in this case, the intensity of Twitter activity bursts or drop-offs). We tested the sensitivity of 2, 5, 10, 20, and 35 km^2 hexagons to determine the minimum size at which the majority of hexagon Twitter count totals were sufficiently large for the Central Limit Theorem to hold ($n>30$), at which the steady state data distributions had minimal kurtosis, and at which minor variations in scale did not have significant impacts on the number of identified extreme bursts or drop-offs.

This research found that data bias was minimized and analytical significance preserved with an optimal study area of 5-15 km^2. Thus, we chose to generate a grid of 10 km^2 polygons. The grid we generated to cover the study regions consisted of approximately 10,000 individual hexagons. Of these, we removed those that did not contain any Twitter data during both the steady and perturbed states to reduce the potential for conflated significance in our results. Our final analysis, excluding Puerto Rico, contained nearly 7,000 10 km^2 hexagons.

2.3.4 Twitter deviation analysis

Upon the completion of our datasets of temporally and spatially segmented Tweet counts, we characterized the distributions of Twitter activity observed for each eight-hour segment for each hexagon across both the steady and perturbed states. Previous research has determined that more original content is produced in areas strongly affected by a hurricane, and, as mentioned, bursts in Twitter activity during a hurricane are significantly correlated with damage (David et al. 2016; Kryvasheyeu et al. 2016). However, even if hurricane-related Tweets peak just after the hurricane peaks, the variability within small geographic areas (in terms of increasing or decreasing activity) has not been explored. We compare the differences in the distributions in Twitter activity counts from steady state to perturbed state to characterize the prevalence of both types of Twitter activity changes: increases, and decreases.

Following our characterization of the Twitter activity in steady and perturbed state, we determined the deviation in Twitter activity from steady to perturbed state on a daily basis for each of the 7,000 hexagon in the study regions. The deviation in Twitter activity for each hexagon was defined using **Equation 1**, where δ_a represents the Twitter activity deviation, a_p represents the number of Tweets (activity) observed in a defined period in a hexagon within the perturbed state, μ_s represents the average number of Tweets observed in that area during the steady state, and σ_s represents the standard deviation of the Tweet counts observed during the steady state. Finally, the deviations for each hexagon were summed across the disparate temporal categories to create daily deviation values for every day of the perturbed state. This process is visualized in **Figure 3**, which shows the hexagonal grid overlaying the Twitter data for a day of Houston, Texas' steady state, the average number of Tweets within each of those hexagons, and the deviation in the number of Twitter posts for one day of the perturbed state.

$$\delta_a = \frac{a_p - \mu_s}{\sigma_s} \tag{1}$$

Finally, we had determined the standardized deviation δ_a in Twitter activity (defined here as the number of original content Twitter posts in a geographic area) from steady state to perturbed state (defined here as the period of time a region is heavily impacted by a hurricane) across a total of more than six weeks (as recommended in [Toepke 2018]) and 70,000 km^2 of the southeastern United States.

With this data, we first characterized the observed distributions of these deviations for each city to identify the relative prevalence of increases and decreases in Twitter activity. Next, to ascertain the relative importance of considering decreases as dangerous, we defined a metric that would rank decreases in activity as relatively unimportant (raw deviation, δ_a) and one that would consider them as equally important as a comparative increase in activity (absolute deviation, $|\delta_a|$). The raw deviation encompasses the direction of the change in activity. When ranked from lowest to highest, the areas that experienced the greatest decrease in activity are lowest ranked, and the areas with the greatest bursts in activity are highest ranked. The absolute deviation encompasses only the magnitude of the change in activity. When ranked from lowest to highest, the areas that exhibited the least change in activity from steady state to perturbed state are lowest ranked, and those that exhibited the greatest amount of positive or negative change are highest ranked.

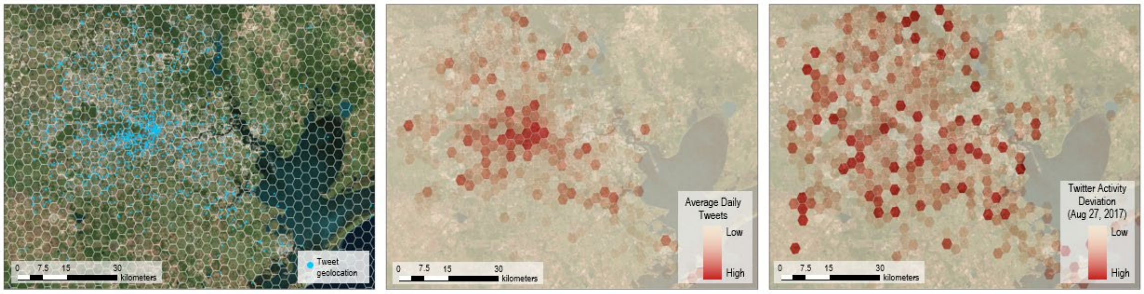

Figure 3. (Left) Houston, Texas overlaid with a white-lined grid composed of 10 km2 hexagons. Blue dots are used to indicate the location of Tweets accumulated during one day of the steady state for the city. (Middle) The average daily Tweet counts for Houston, Texas across the steady state period. Darker red hexagons indicate higher average daily Tweet counts, while paler yellow hexagons indicate lower average daily Tweet counts. (Right) The absolute deviations in Twitter activity for Houston, Texas on August 27th, 2017, Harvey's second landfall. Darker red hexagons indicate higher amounts of change in Twitter activity from steady state averages, and paler yellow hexagons indicate lower amounts of change.

2.3.5 Infrastructure damage validation

Having defined the deviation metrics for analysis and gathered and segmented the Twitter data, we needed to assess the correlation between each metric and the hurricane damage that social media crisis applications seek to identify. To validate the relevance of Twitter activity deviation, we used FEMA Building Level Damage Assessments (referred to as FEMA Damage Assessments) as an infrastructural damage metric. The FEMA Damage Assessments were created as soon as was safely possible following the hurricane and are available as point shapefiles that represent buildings that have been classified as having "No Damage", being "Affected", having "Minor Damage", having "Major Damage", and being "Destroyed" (The Federal Emergency Management Agency 2016). We reinterpreted this range into a continuous numeric scale from 0, representing "No Damage", to 4, representing "Destroyed". This more subjective scale would represent the overall potential danger to a building's inhabitants during the hurricane more judiciously than a monetized scale, which might amplify damage to less vulnerable communities (Villegas et al. 2018).

We used ArcGIS to determine a variety of statistics for the damage ratings assigned to buildings within each hexagon, including the count, average, and maximum. The FEMA Damage Assessment point data tended to be present in clusters of varying sizes across the cities. On further investigation, we found these clusters to have peak amounts of infrastructure damage at their centers and varying amounts of less-damaged infrastructure in the surrounding areas. For instance, in the immediate area surrounding each cluster of "Destroyed" buildings, there would be anywhere from five to one hundred buildings rated as having "Minor Damage" or having been "Affected". Because of this, some areas with the same number of destroyed buildings would have substantially different average damage ratings, most often due to differences in neighborhood size. This would be a confounding factor in using the average. Also, as this research is primarily concerned with people in life-threatening danger, we determined that preserving the presence of "Destroyed" buildings in the infrastructural damage metric would be critical to the utilization of our findings. Thus, to mitigate the confounding effect of neighborhood size and to preserve the magnitude of extreme situations, we used the maximum damage rating within each hexagon instead of the average for our damage metric.

2.3.6 Correlations between Twitter activity and damage

To compare the relative correlations between each deviation metric and the infrastructural damage metric identified above, we utilized both Kendall and Spearman rank correlation tests (Kendall 1938; Sedgwick 2014). These two tests compare the similarities in the order of the hexagons when ranked from lowest to highest deviation and when ranked from lowest to highest infrastructural damage. For the raw deviation, the hexagons were ordered from highest decrease in activity to highest increase in activity; for the absolute deviation, the hexagons were ordered from least amount of change in activity to greatest amount of activity change. It should be noted that we chose rank correlation tests because the FEMA Damage Assessment scale is both ordinal and non-continuous. Although there is a clear underlying variable that has been discretized, it would not be accurate to assume that the values are evenly spaced with respect to damage impact. For the final analysis, we determined the Kendall rank correlation coefficient τ and the Spearman rank correlation coefficient ρ for both the raw and absolute Twitter activity deviation for each day of the perturbed state for each city.

2.4 Results

2.4.1 Twitter Usage Characterization

As described above, we first characterized the distributions of the number of Twitter posts for each eight-hour temporal segment (i.e., the Twitter activity within that period) for the steady states and the perturbed states. A violin plot of that activity, comparing the distributions of Twitter activity for both states for each city, is presented as **Figure 4**. Most of the distributions of steady state Twitter activity per eight-hour period appear to be bimodal, and the perturbed state distributions are more likely to have a single local maxima. Additionally, the perturbed state average and median values appear to be slightly lower than their steady state counterparts. It should be noted that the violin plots show only the central 95% of the data. As such, although the perturbed state values appear to have a significantly smaller range than the steady state values, the more extreme perturbed state values were more likely to be categorized as outliers and not included. We still find that the range of Twitter counts is larger for the steady state than the perturbed state, reinforcing the need for a robust amount of data to define a steady state. This figure further demonstrates the differences in the number of Tweets produced per eight-hour period within each city: Houston, Texas and Miami, Florida consistently produce more than one thousand Tweets in an average eight-hour period, while a significant number of areas in both Biloxi, Mississippi and Beaumont, Texas produce fewer than ten.

Following the recognition that the average Twitter activities of the perturbed state appeared to be generally lower than those of the steady state, we characterized the density of raw deviations in Twitter activity from the steady state to the perturbed state. The distribution of the deviations in Twitter activity observed for each hexagon for each day of the perturbed state are shown in **Figure 5**. The distributions of Twitter deviation have similar characteristics across the study regions. Each distribution is unimodal with a peak in Twitter activity deviations at approximately -0.2, a steep drop-off in density at 0, and a long, thin right tail. The long right tail is expected because the distribution is bounded to the left; an area's activity level can only fall to zero, so each area's deviation is bounded by its steady state average. On review of the data, the effect of the long right tail is especially evident for Houston, where the total number of Tweets peaks on August 26th and 27th; however, these two days also show the highest number of areas exhibiting overall decreases from the norm.

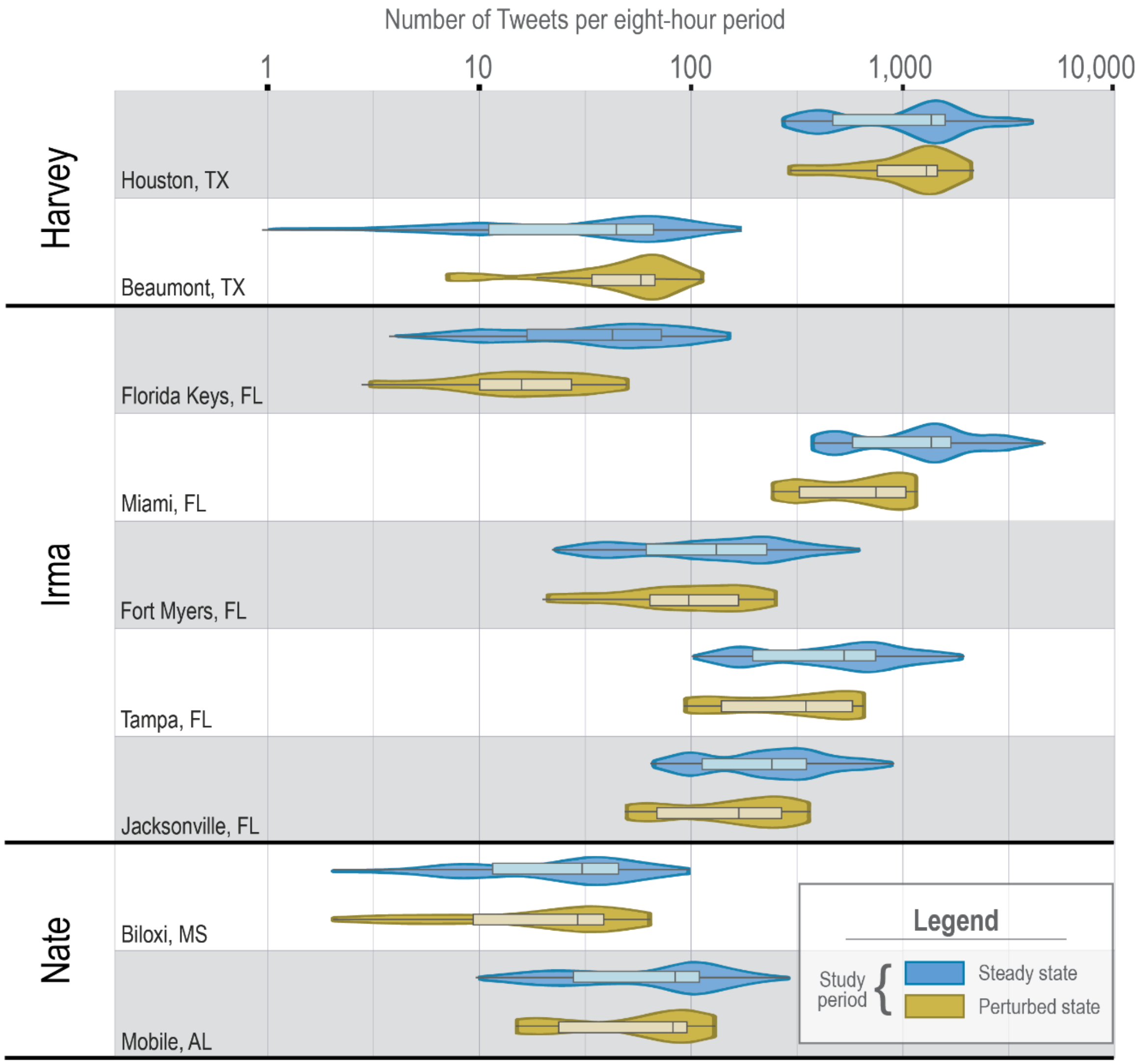

Figure 4. Violin plot depicting the distributions of Twitter activity counts recorded across all of the eight-hour periods during the steady and perturbed states for each region of interest. The distributions of the steady states are depicted in blue, and the distributions of the perturbed state are depicted in yellow. The width of each violin depicts a greater density of values within that region, and the length of the violin represents the range of 95% of the data. Box plots are overlaid on each violin that depict the median as a black line down the center and one standard deviation on either side of the median composing the first and third quartile.

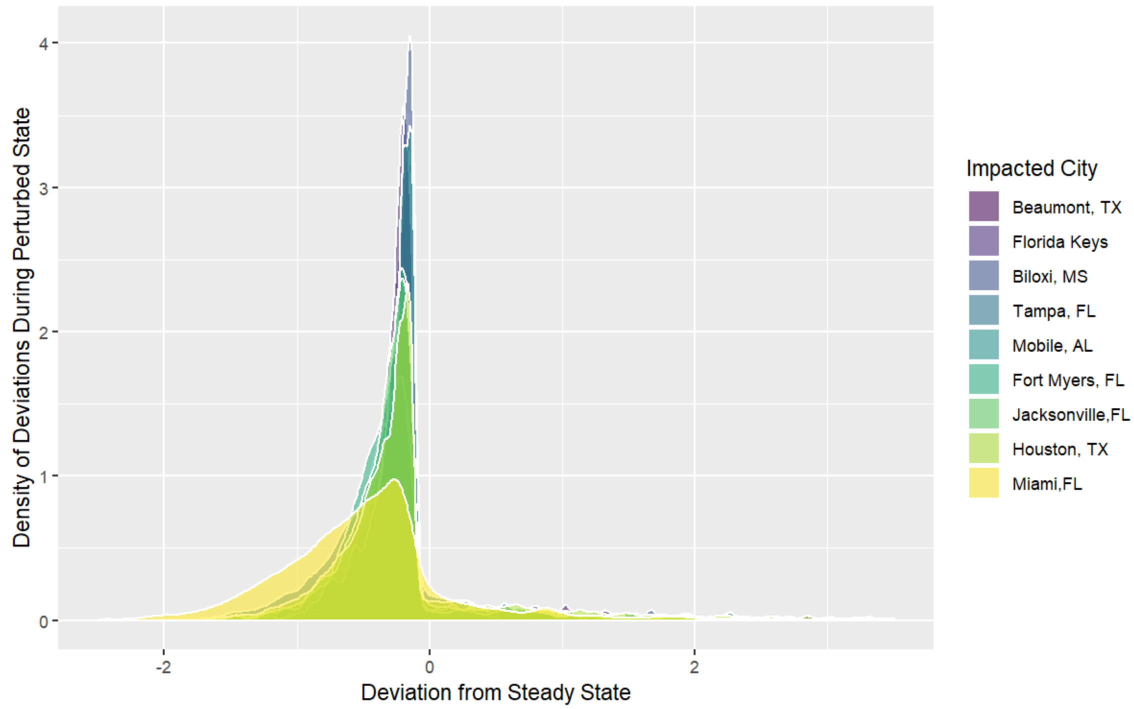

Figure 5. Density plot depicting the distribution of deviation values recorded from all of the hexagons across all of the perturbed state (seven days) for each city. Negative values indicate hexagons that decreased in activity from the steady state to the perturbed state. The different colored areas indicate the different cities.

The regions of negative deviations show higher overall densities, although the positive deviation regions have greater range. The lowest observed deviation in Twitter activity was -2.7, and the highest was 81.2. Miami shows the greatest spread, and Beaumont shows the least. This shows that very few regions did not change at all, the majority of areas decreased slightly in Twitter activity in the perturbed state, and a few outlier areas increased a drastic amount.

2.4.2 Rank comparisons

Following our characterization of the observed Twitter activity, we derived the Spearman and Kendall rank correlation coefficients for each day of the perturbed states for

1) the absolute Twitter activity deviation and the maximum recorded infrastructure damage within the hexagons, and 2) the raw Twitter activity deviation and the maximum recorded infrastructure damage within the hexagons. These values are presented in **Table 2** and **Table 3**. We first focused on the correlation coefficients derived for the day on which the respective cities experienced the greatest amount of damage. For most cities, this day is landfall; however, Hurricane Harvey had a second landfall on August 27th that did more damage through flooding than its initial impact on August 25th. To reflect this, we use August 27th as the principal day of analysis for both Houston, Texas and Beaumont, Texas. The rank correlation coefficients for the deviations determined on the day of maximum damage are compared for each city in **Figure 6**.

Spearman and Kendall rank correlation coefficients are shown by the circle markers at the end of each line of the lollipop chart, marked with an “S” or “K” respectively. Blue markers and lines indicate the result of the rank correlation test between infrastructural damage and the absolute Twitter activity deviation, and yellow markers and lines indicate the result for the correlation between infrastructural damage and the raw Twitter activity deviation. Hollow markers indicate that the derived rank correlation coefficient for that test was statistically insignificant ($p>0.05$). The cities are ordered chronologically by hurricane and time of landfall.

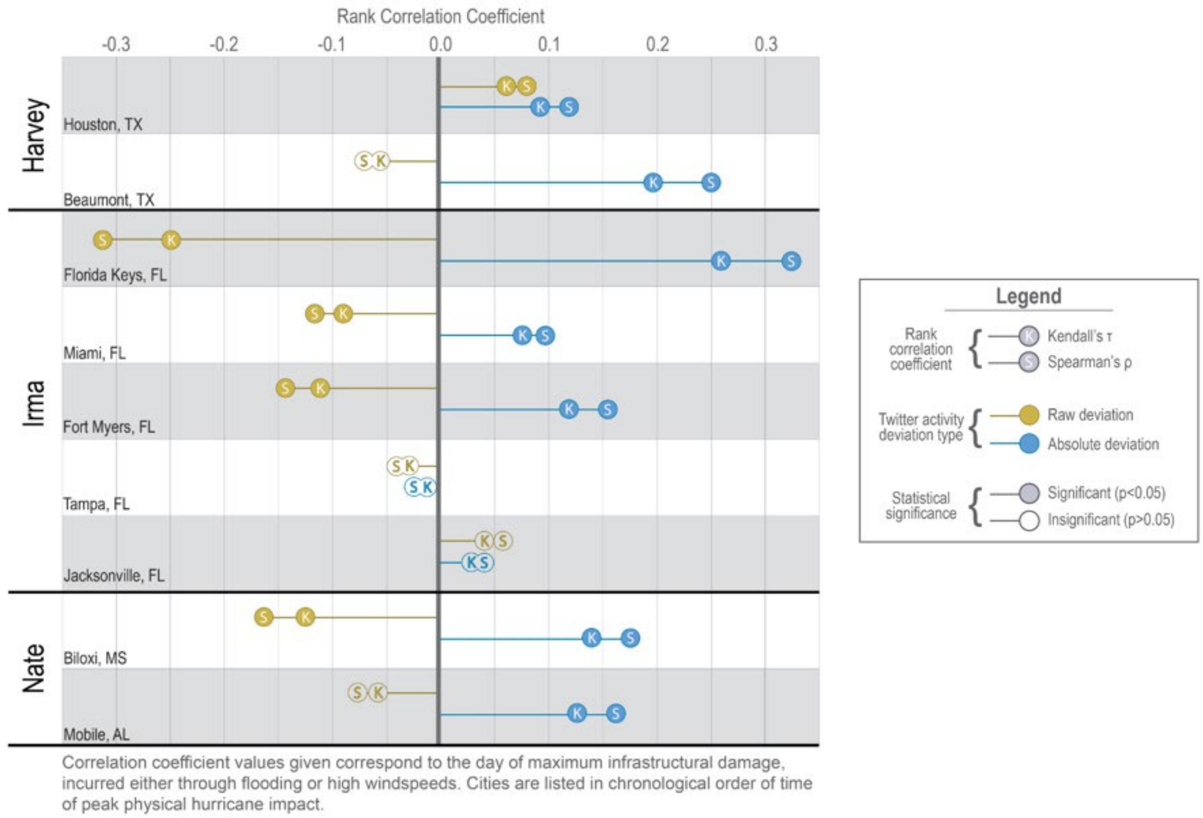

Figure 6. Lollipop chart comparing the rank correlation coefficients for the raw and absolute Twitter deviations across study regions on the day the regions experienced the most infrastructure damage according to recorded rainfall and wind conditions and National Oceanic and Atmospheric Administration (NOAA) reports. Kendall's τ is depicted with a circle containing a "K" and Spearman's ρ is depicted with a circle containing an "S". The values for the raw deviation depicted in yellow, and those for the absolute deviation are depicted in blue. Values that were statistically significant ($p<0.05$) are shown with filled circles, while those that were statistically insignificant ($p>0.05$) are shown with hollow circles.

Table 2. Spearman rank correlation coefficients between social media deviation metrics and infrastructure damage across the perturbed state.

Spearman's ρ																		
Day	Harvey				Irma										Nate			
	Houston		Beaumont		Florida Keys		Miami		Fort Myers		Tampa		Jacksonville		Biloxi		Mobile	
	Absolute	Raw	Absolute	Raw	Absolute	Raw	Absolute	Raw	Absolute	Raw	Absolute	Raw	Absolute	Raw	Absolute	Raw	Absolute	Raw
Day -1	0.1*	-0.06*	0	0.03	0.39*	-0.26*	0.07	-0.1*	0.14*	-0.17*	-0.01	-0.09*	0.02	-0.04	0.08	-0.09	0.2*	-0.1*
Day 0	0.07*	-0.05*	0	0.06	0.32*	-0.31*	0.1*	-0.12*	0.15*	-0.14*	-0.02	-0.04	0.04	0.06	0.18*	-0.17*	0.16*	-0.08
Day 1	0.07*	-0.05*	0.05	0.01	0.37*	-0.29*	0.13*	-0.12*	0.16*	-0.22*	0.01	-0.05	0.04	0.02	0.13*	-0.02	0.08	-0.05
Day 2	0.12*	0.08*	0.25*	-0.07	0.35*	-0.23*	0.06	-0.07	0.13*	-0.18*	0.03	-0.09*	0.03	0.01	0.16*	-0.11	0.16*	-0.08
Day 3	0.1*	0.06*	0.07	0.01	0.38*	-0.28*	0.13*	-0.12*	0.16*	-0.21*	-0.01	-0.06*	0.06	-0.08*	0.15*	-0.11	0.14*	-0.08
Day 4	0.12*	0.05*	0.07	0.09	0.36*	-0.29*	0.08*	-0.08*	0.12*	-0.24*	0.04	-0.08*	0.06	-0.04	0.15*	-0.02	0.15*	-0.03
Day 5	0.11*	-0.03	0.04	0.09	0.38*	-0.29*	0.11*	-0.11*	0.13*	-0.15*	0.03	-0.06	0.03	-0.08	0.08	0	0.15*	-0.07
Day 6	0.12*	-0.02	0.09	0.11*	0.42*	-0.35*	0.13*	-0.11*	0.11*	-0.12*	0.01	-0.03	0.04	0.01	0.09	-0.17*	0.13*	-0.05

* indicates $p < 0.05$

Table 3. Kendall rank correlation coefficients between social media deviation metrics and infrastructure damage across the perturbed state.

Kendall's τ																		
Day	Harvey				Irma										Nate			
	Houston		Beaumont		Florida Keys				Houston		Beaumont		Florida Keys				Houston	
	Absolute	Raw	Absolute	Raw	Absolute	Raw	Absolute	Raw	Absolute	Raw	Absolute	Raw	Absolute	Raw	Absolute	Raw	Absolute	Raw
Day -1	0.08*	-0.04*	0	0.03	0.31*	-0.21*	0.06	-0.08*	0.1*	-0.13*	-0.01	-0.07*	0.02	-0.03	0.06	-0.07	0.15*	-0.08*
Day 0	0.05*	-0.04*	0	0.05	0.26*	-0.25*	0.07*	-0.09*	0.12*	-0.11*	-0.01	-0.03	0.03	0.04	0.14*	-0.13*	0.13*	-0.06
Day 1	0.05*	-0.03*	0.04	0.01	0.3*	-0.23*	0.1*	-0.09*	0.12*	-0.17*	0.01	-0.04	0.03	0.02	0.1*	-0.02	0.06	-0.04
Day 2	0.09*	0.06*	0.2*	-0.06	0.28*	-0.19*	0.05	-0.06	0.1*	-0.14*	0.03	-0.07*	0.03	0.01	0.12*	-0.08	0.13*	-0.06
Day 3	0.08*	0.04*	0.05	0.01	0.31*	-0.23*	0.1*	-0.1*	0.13*	-0.17*	-0.01	-0.05*	0.05	-0.06*	0.11*	-0.08	0.11*	-0.07
Day 4	0.09*	0.04*	0.05	0.07	0.29*	-0.24*	0.06*	-0.06*	0.09*	-0.18*	0.03	-0.06*	0.05	-0.03	0.11*	-0.02	0.12*	-0.02
Day 5	0.08*	-0.02	0.03	0.07	0.31*	-0.23*	0.09*	-0.09*	0.11*	-0.12*	0.02	-0.05	0.02	-0.06	0.06	0	0.11*	-0.05
Day 6	0.09*	-0.02	0.07	0.08*	0.34*	-0.28*	0.1*	-0.09*	0.08*	-0.09*	0.01	-0.02	0.03	0.01	0.07	-0.13*	0.1*	-0.04

* indicates $p < 0.05$

The data show that, when the cities are hit by peak infrastructural damage, the absolute Twitter activity deviation in Twitter activity consistently has a stronger, significant correlation with that infrastructural damage than the raw deviation. Seven of the nine impacted cities show a significant correlation between the absolute Twitter activity deviation and the amount of destruction wrought by the hurricane, and the other two cities do not exhibit a significant relationship between either the raw or absolute Twitter activity deviation. Additionally, for Hurricane Irma and Hurricane Nate, the strength of the correlation between absolute Twitter activity and infrastructural damage decreases with increasing geographic and temporal distance from each storm's first landfall. This is particularly evident in the strength of the correlation for the Florida Keys, hit first and hardest, as compared to that of Tampa and Jacksonville, which were hit a full day later and are more inland.

Additionally, we see an inverse relationship between the absolute and raw Twitter activity deviations for many of the cities. The correlation strengths are not perfectly inverse, although, as would be expected, the absolute deviation tends to be much stronger when the difference between the absolute and raw is greater. This is particularly true for cities impacted by Hurricane Irma. Houston, Texas, Tampa, Florida, and Jacksonville, Florida are the only cities for which the two deviation (raw and absolute) metrics are directly related, and the correlations for both sets of metrics are statistically insignificant for the latter two cities. The two cities for which the correlation between infrastructure damage and absolute Twitter activity deviation is significant while the raw deviation is not are Mobile, Alabama and Beaumont, Texas: both comparatively smaller cities.

Looking beyond landfall towards urban recovery, we turned our attention to the duration of the correlations' significance. In comparing the correlations across time, we analyzed the rank correlation coefficients of the absolute and raw Twitter activity coefficients with infrastructure damage for each day of the week following landfall. The results of that analysis for Houston, Texas are shown in **Figure 7**.

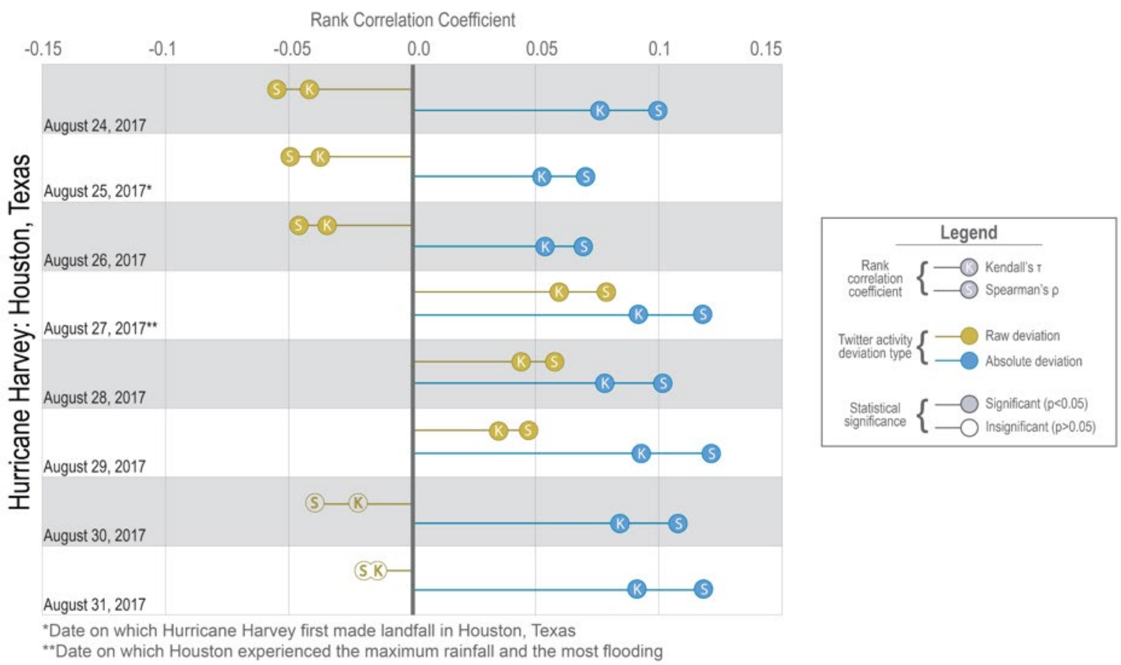

Figure 7. Lollipop chart comparing the rank correlation coefficients for the raw and absolute Twitter deviations observed in Houston, Texas across the days of the perturbed state. Kendall's τ is depicted with a circle containing a "K" and Spearman's ρ is depicted with a circle containing an "S". The values for the raw deviation depicted in yellow, and those for the absolute deviation are depicted in blue. Values that were statistically significant ($p<0.05$) are shown with filled circles, while those that were statistically insignificant ($p>0.05$) are shown with hollow circles.

As the trends are not uniform across each of the cities, we have divided our description of the results by the hurricane that impacted each city.

2.4.3 Hurricane Harvey (Aug 24-Aug 31)

Similar to the results observed by Kryvasheyeu et al., there is a weak but significant correlation between incoming infrastructural damage and Twitter usage behaviors as early as the day before landfall, although not prior to that (Kryvasheyeu et al. 2016). This correlation decreases on the date of landfall, then rises for both metrics on the day of greatest observed damage. A key difference between the raw and absolute deviation correlation coefficients, however, is that the raw deviation drops following its peak on the day of Harvey's second landfall, and the absolute deviation maintains both its strength and its statistical significance for five days following its peak.

For Beaumont, however, instead of a weak to moderate correlation strength for the absolute deviation in Twitter activity across the duration of the perturbed period, the only significant correlation was one of moderate strength on the day of peak infrastructural damage. All other days had insignificant correlations, and all correlations with the raw Twitter activity deviation were insignificant.

2.4.4 Hurricane Irma (Sept 9-Sept 17)

We observed similar short-term trends in the cities affected by Hurricane Irma. The Florida Keys' absolute Twitter activity deviation correlation is the strongest by far among the cities, and it is significant and moderately strong across the duration of the perturbed state. This contrasts with Miami and Fort Myers, which have correlation strengths that peak after landfall and then slowly decrease in strength. The correlation between raw Twitter activity deviation and infrastructure damage is negative for each of these three cities and is not significant across the duration of the perturbed state following landfall. The two northernmost cities, Tampa and Jacksonville, have insignificant correlation coefficients across much of the perturbed state. Jacksonville exhibits insignificant correlations for both raw and absolute Twitter activity deviations; however, in direct contrast to the other cities, Tampa exhibits significant correlations only for raw Twitter activity deviations.

2.4.5 Hurricane Nate (Oct 8-Oct 15)

The absolute Twitter activity deviation observed in Biloxi, Mississippi and Mobile, Alabama are significantly correlated with infrastructure damage for landfall and for at least four of the days following landfall. The absolute Twitter deviations exhibit correlation strengths that peak on landfall and then slowly decrease in strength. The correlation strengths of the two cities are fairly similar to each other and follow similar trends across the perturbed state.

2.5 Discussion

First, in our characterization of Twitter activity changes, we see that Twitter activity during a hurricane decreases in most local geographic areas. This indicates the presence of Simpson's Paradox within the data: a few super-users Tweeting excessively result in an overall increase in Twitter activity, despite more individual areas showing a decrease in activity. Because we see non-uniform/non-random changes in Twitter activity, these results additionally show that we cannot universally characterize an area's "Tweeting population" (number of users who are active Tweeters) *during* a crisis by its Tweeting population *out of* crisis. Our data reinforces the necessity of using a variety of historical contextualization metrics to define the applicable scope of social media data in crisis applications. As has been called for in the literature (Chen et al. 2013) and, as we have shown, utilizing historical Twitter activity data to develop a steady state is one worthwhile contextualization. We are working with less data from most locales during a crisis, and we need to factor that into our data processing. Exactly why less data is produced and the significance of geographic clusters of activity increases and decreases still need to be explored.

In testing the importance of this missing data through our methodology, we show that seven of nine cities showed that the absolute deviation metric had a stronger, more positive, statistically significant correlation with damage as compared to the raw deviation on the day of maximum experienced damage. This soundly confirms **H1**: sharp social media decreases are more likely to be signals for increased infrastructural damage than an absence of danger. Absence of evidence is not evidence of absence; indeed, in this case, the opposite is more likely.

It should be noted that the correlation strength values we identify are slightly lower than those found in previous literature (0.1-0.35 instead of 0.2-0.65, as in Kryvaysheu et al. [2016]). The decreased strength could be due to the decreased geographic scales of analysis; the eliminated confounding (and potentially strengthening) influence of using ZCTAs; or the method of standardizing the data to a steady state. As we focus on the direct comparison between the absolute ($|d_a|$) and raw (d_a) Twitter activity deviations to identify the *relative* importance of drop-offs in danger identification, we believe our conclusions still hold strong.

As described above, seven of nine cities confirmed **H1** for data produced on the day of maximum damage; six cities showed this trend to be consistent throughout the duration of the perturbed state. Of the three cities which did not, two of them showed no significant correlation between either raw or absolute Twitter activity deviation on more than one day of the perturbed state. These results are likely a combination of variances in the demographic composition of the cities themselves and the strengths of the hurricanes. For instance, Houston and Beaumont, Texas are geographically close; Harvey struck both with a similar strength. However, Beaumont had fewer “destroyed” FEMA Damage Assessments and fewer FEMA Damage Assessments overall per household (U.S. Census Bureau 2016). This could be because Houston was drowning beneath the overflow from Buffalo Bayou (Nyaupane et al. 2018), and Beaumont—which is tangential to an inlet water body instead of surrounding one—was not so heavily inundated. Beaumont also had one of the lowest steady state Twitter distributions, which may have contributed substantially to the lack of significant correlations. From the comparison of these two cities, we theorize that there is a direct relationship between hurricane damage and the relative strength of our “silence” metric. This is also echoed in literature on Hurricane Sandy: the area of New York that had the most deaths was also relatively social media silent (Shelton et al. 2014). If social media silence becomes more significant with increasing hurricane damage, and if hurricane damage is expected to increase (Gall et al. 2011), then this implies an increasing need to consider the impacts of the silence we identify.

In considering the impacts, we need to consider cause and prevention. Which of the drop-offs are due to failures in urban resilience, and which due solely to intentional human behavior choices? Interrupted network access could have been caused by power outages combined with a lack of access through a paid mobile network or through lack of access to external networks. It should be noted that many lower socioeconomic groups rely on internet access through external networks, such as Starbucks, instead of at-home networks (Khan et al. 2016). As noted by Zou et al., vulnerable populations contribute comparatively less to social media streams than less vulnerable populations during a disaster (Zou et al. 2018). A disparity in internet access during a hurricane could contribute to a disparity in the ability of vulnerable populations to Tweet and thus inequity in the un-contextualized usage of social media data in crisis applications. Our research currently cannot show whether a decrease in Twitter activity is due to specific, intentional posting choices or infrastructure failures, and the subject should be explored in future work. All posting intentions equal, if higher socioeconomic have a higher likelihood of maintaining internet access than lower socioeconomic groups, and if crisis applications contribute more resources to regions on the premise of Tweet bursts or key word identification, vulnerable populations may receive an inequitably low proportion of social media-directed resources.

As the crisis informatics community continues to evolve, and as emergency responders are increasingly monitoring social media during a crisis (Murthy and Gross 2017), this consideration of how well the existing data represents different populations will be critical to equity. Areas with higher vulnerability scores have already been shown to be more poorly served by existing emergency response services like hurricane evacuation (Bian and Wilmot 2017); are we making this disparity worse? The existing methods advising social media tools are currently blind to the needs of people unable to Tweet during a disaster due to immediate personal danger or failures in energy infrastructure. The possible oversight of these populations advocates the need for social media analyses that are capable of detecting their need. Particularly as recent research into vulnerable populations has shown that the poorest populations—and poorest countries—are going to be hit first and hardest by the effects of climate change (Schiermeier 2018), the way we process information needs to be at least peripherally aware of the differences in digital capabilities between the poor and the affluent; the areas of a city with poorly maintained infrastructure and those with resilient buildings; and those residents prepared for an incoming disaster and those taken entirely unawares.

2.5.1 Limitations and Future Directions

With respect to our method of using Twitter activity deviation instead of spikes, a potential flaw is that the decrease in activity primarily indicates areas that people have evacuated, or commercial sectors of the city that have closed in anticipation of danger. We attempted to alleviate this risk by removing hexagons without Tweets in both the steady and perturbed states. On review of the spatial correlation of our data, the distribution of hexagons that increased and decreased in Twitter activity does not appear to be consistent with county evacuation zones. With respect to the problem of reduced mobility to commercial sectors, our use of a scalable hexagonal grid should minimize the partitioning of our study area by residential and commercial sectors that would have been a factor in using ZCTAs. Although evacuation is surely a factor in some of the wide-scale decreases in Twitter activity, we note that neither Houston nor Beaumont, TX issued comprehensive evacuation orders. A mobility study using cell phone data identified an approximate 5% decrease in unique users on the Texas coast (Marzuoli and Liu 2019); even with this minor decrease, both cities match our conclusions. Of course, future research needs to consider how and where to apply the described “silence” metric to account for both silence and evacuation. The effect of broad-scale, enforced evacuation on the deviation correlation should be identified and accounted for.

Tangential to that, we note that very little true information was produced through Twitter during Hurricane Maria on Puerto Rico; the damage threshold at which the form of humans-as-sensors data changes is important to consider. With an area in complete social media silence, little can be done with our approach. Multiple datastreams and data approaches will be necessary to increase the robustness of our crisis informatics applications.

From an analytical perspective, our usage of 10 km^2 hexagons was primarily based on summary statistics, recommendations from the field of critical GIS, and experience with the detriments of the modifiable areal unit problem in census tracts (Jelinski and Wu 1996; Saib et al. 2014). Future research needs to focus on determining at what spatial and temporal scales the identified relationship is strongest, and at which spatiotemporal scales the relationship breaks down.

Future work will broadly need to determine when our "silence" metric should be more rigorously utilized. Our research showed that there were several hurricanes for which it was an indisputably stronger metric for damage, and two for which it was entirely insignificant. As stated earlier, we need more research to understand if the presence or absence of "silence" is linked to disaster severity, infrastructural resilience, disaster preparedness, vulnerable populations, or something else entirely. Following that more theoretical work, additional research should also explore how to incorporate "silence" from a steady state into new and existing social media crisis applications.

2.6 Conclusion

Our initial characterization showed the existence of sample bias created when a region transitions from a normal state to a crisis state. Following that, in testing **H1**, we show that the sample bias introduced by this transition is significant, and that failing to recognize social media drop-offs could put people—possibly the most vulnerable people—at risk. By analyzing social media activity deviations and recognizing social media drop-offs, we show that a metric that prioritizes drop-offs is more highly correlated with infrastructure damage than one that ignores them. Not every social media user can or will contribute to the humans-as-sensors datastream during a crisis, and, as is explicitly shown here, the social media data shadows generated by this behavior could hide disaster.

Crisis informatics is continually developing new methods of incorporating social media into crisis response, and we are applying it to everything from event detection to resource distributions. However, one weakness of these methods is the exclusive focus on only the data produced during a crisis situation. These methods neglect to consider populations that are suddenly unwilling or unable to Tweet and have dropped out of the social media data stream. Within this paper, we test for the prevalence of social media drop-offs and test whether sudden social media silence in an area could be an indicator of a crisis situation. We find that most geographic areas, when assessed at uniform size and shape, decrease in Twitter activity during a hurricane. For the latter, we find that greater Twitter activity drop-offs are significantly correlated with *more* infrastructure damage, not less. The importance of recognizing and including social media "silence" is consistent across seven of nine cities impacted during the 2017 Atlantic Hurricane Season.

Ultimately, our research shows that social media data shadows are more likely to be hiding people in danger, not less. To use humans-as-sensors data in the most equitable and actionable way, crisis researchers need to consider both the sound *and* silence of the Tweets.

2.7 Acknowledgements

This material is based upon work supported by the National Science Foundation under Grant No. 1760645. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

CHAPTER 3. TIPPING THE SCALES: HOW ANALYTICAL SCALE AFFECTS THE INTERPRETATION OF SOCIAL MEDIA BEHAVIOR IN CRISIS RESEARCH

3.1 Abstract

Our relationship with technology is constantly evolving, and how we use technology in disasters has evolved even faster. Understanding how to utilize human interactions with technology and the limitations of those interactions will be a crucial building block to contextualizing crisis data. The impact of geographic scale on behavioral change analyses is an unexplored facet of our ability to identify relative severities of crisis situations, magnitudes of localized crises, and total durations of disaster impacts. Within this paper, we aggregate Twitter and hurricane damage data across a wide range of geographic scales and assess the impact of increasing scale on both the recognition of extreme behaviors and the correlation between activity and damage. The power-law relationships identified between many of these variables indicates a direct, definable scalar dependence of social media aggregation analyses, and these relationships can be used to inform more intelligent, equitable, and actionable social media usage in emergency response.

3.2 Background

As the supply of data from humans-as-sensors continues to increase, understanding individual data streams in the context of our multi-layered and multi-networked society is becoming more difficult. Social media is increasingly looked to as a potential source of additional information in the notoriously information-scarce environment of crises (Reuter and Kaufhold 2018). The crisis informatics field has continued to flourish and expand alongside the seemingly ever-increasing quantities of available social media data and methods of analyzing that data. As a result, the applications for social media during crises has expanded to include event detection (Sakaki et al. 2010), resource availability and need (Choe et al. 2017), and mobility monitoring (Wang and Taylor 2016a). Analytical methods range from sieving individual posts for information (Ashktorab et al. 2014), to analyzing geographic changes in sentiment and behavior for informing gestalt-level decisions (Jongman et al. 2015; Kryvasheyeu et al. 2016).

The big data revolution has clearly opened a vast area of possibilities for crisis response (Qadir et al. 2016). One of the greatest strengths of the field is the diversity it contains, and the range of techniques available for the processing of this ever-increasing and changing pool of data. Applications from the field are being used by international aid organizations (Imran et al. 2014), and strides are being made for local response implementations of social media analysis as well (Tapia and Moore 2014). That said, the range of diversity of applications and methods can also impede the process of building a solid foundation. Researchers both external and internal to crisis informatics have noted criticism of social media applications' limits with respect to data bias, social inequality, and lack of confirmed validity (Imran et al. 2015; Jiang 2018).

More data is available; however, big data is not complete data. There has been a consistent call for us to critically interrogate the assumptions and capabilities of big data in the context of our political and urban usage (Boyd and Crawford 2012). As the reach and amount of available data increase, holes in that data become both less obvious due to the existing volume and yet more harmful due to the increasing prevalence of that data's use (Morstatter and Liu 2017). Social media, especially when used for crisis response, is not exempt from this call. This is especially true in the case of crisis response, where information availability alone can tip the scales of resource distribution. To ensure more equitable and intelligent use of social media data in crisis response, researchers need to understand the social, spatial, and sociospatial limitations of that data. One critical piece of that understanding is understanding the geographic scale at which social media data is capable of identifying disasters, and how much information is gained or lost by varying that geographic scale.

Geographic scale is less important in analyses sifting through individual posts, but it becomes more relevant when determining the likelihood of each of those posts appearing in a specific place and time. Spatiotemporally-aggregated data can be key to identifying an expected baseline level of activity (Toepke 2018a); identifying the proportions of a population that Tweet (Mislove et al. 2011); identifying drop-offs in activity alongside spikes in activity (Samuels et al. 2018a); and ultimately providing a social lens through which the social benefits or ramifications can be, at a minimum, glimpsed (Shelton et al. 2015).

Aggregate behaviors in real-time can be used to analyze relative disaster severity and magnitude. As is often echoed, disasters are not disasters because of high wind speeds or unprecedented amounts of flooding; disaster are disasters because of how they interrupt and, sometimes, forever change a society's functions. Disasters are inherently social and, ultimately, defined by societal vulnerabilities. The impacts and vulnerabilities associated with a disaster need to be defined in multiple dimensions, with an emphasis on social norm disruptions (Guan and Chen 2014). Spatiotemporal aggregation, then, can give us the pre- and post-impact phase reference points called for by Guan and Chen (2014).

However, we currently lack understanding of the scale at which to generate these reference points. Previous research, particularly focusing on social disruptions, has identified the existence of a scalar dependence on the correlation between hurricane damage and Twitter activity fluctuations (Shelton et al. 2014), but did not further investigate it. Subsequent calls for further research highlight how understanding the scalar dependencies of social media data will improve our reference for the data's place in geographic, temporal, and social space (Jessop et al. 2008), thus improving our total understanding of the social significance of Twitter activity trends.

Lastly, a massive disaster such as a hurricane or an earthquake is composed of hundreds of small ones: flooded neighborhoods, downed overhead power transmission lines, and trees thrown through roofs by gale-force winds (Wurman and Kosiba 2018). These disasters happen to more than individuals, but less than the whole of society. Disasters of varying magnitude can happen to small neighborhoods, along vast swathes of a river, or through power outages across a city. If a 911 call can recognize a disaster happening to an individual, at what scale can social media recognize and triage emergencies affecting more than individuals? Is it limited to disasters occurring to thousands of people, or can it also identify disasters affecting smaller groups? The more we understand the capacity of humans-as-sensors to identify the location, relative severity, and magnitude of the localized disaster, the more useful social media will be to emergency response (Raue et al. 2013). Understanding how scale impacts the recognition of behavior will also enable us to reduce the obfuscation of any minority behaviors occurring at small scales that are drowned out by those happening at larger ones (Chen et al. 2013).

Within this paper, we analyze and present the scalar dependencies of aggregate social media analyses. We focus primarily on the ability of social media to identify localized disasters, i.e., to distinguish groups or areas that are being impacted by the disaster more extremely than the broad geographic region. The connection between social media activity and the presence of extreme danger or disaster has been noted in several pieces of literature (Guan and Chen 2014; Kryvasheyeu et al. 2015), and we are specifically testing the scalar dependencies of that connection. In order to do this, we test at different scales 1) the ability of Twitter to identify distinct clusters of similar behavior, 2) how much behavior is identified as non-normal or extreme during a disaster, and 3) how the strength of the correlation between Twitter activity and hurricane damage varies. Understanding the shape of the relationships between these three factors and changing scale will improve our understanding of how to maximize the benefits of decreasing scale (more specificity of place) while minimizing the costs (less confidence in correlations). These assessments are codified in the following hypotheses:

H1. The distribution of changes in social media behavior and the identification of behavioral clusters is statistically different at smaller scales.

In order to assess this hypothesis, we qualitatively investigated the statistical distribution of Twitter representation (Tweets per person) across the Houston Metropolitan Area using each of the spatial nets. We secondarily quantitatively investigated the scale at which the distributions of Twitter representation cannot be statistically distinguished between increasing spatial aggregation scales for both steady and perturbed state days.

Provided that some aspects of Twitter behavior are scale dependent, we should be able to use the dependency to identify how scale affects crisis-relevant analyses. Most of social media crisis analyses operate on the assumption that humans affected by crisis will change their behavior, and increasing amounts of evidence show that hurricanes produce localized crises at a small scale (Lieberman-Cribbin et al. 2017; Wurman and Kosiba 2018). In order to be useful to crisis managers who need specific and localized information, we should understand the smallest scale at which in-crisis behavior changes are identifiable, and the potential information value trade-offs of decreasing or increasing the scale of analysis. This leads to our second hypothesis:

H2. The identification of crisis-induced, extremely high or extremely low amounts of non-normal Twitter activity is scale-dependent.

For this second hypothesis, we quantitatively assess the percentages of areas that have deviated from their steady state norms during Hurricane Harvey. The assumption of crisis informatics is that disasters break the ability of society to function normally (Guan and Chen 2014). How social media identifies and codifies those breaks in normal functioning is important to understand for trying to identify small-scale crises in a sea of larger crises. We also need to understand if the aforementioned connection between social media activity and hurricane damage is more or less consistent for identifying those small-scale crises. This leads us to our third hypothesis:

H3. The strength of the correlation between Twitter activity and infrastructural damage is scale-dependent.

For this, we test previously-identified correlations between Twitter activity deviations and records of hurricane damage at increasing spatial scales. Knowing how scale affects correlation strength, and therefore affects how confident we can be that activity is indicative of a disaster, is important for communicating with emergency managers.

These quantitative comparisons of the representative capabilities of aggregation techniques will also inform future tools and algorithms that seek a real-time metric for human need expressed through social media. To address these hypotheses, we chose to focus our efforts on the city of Houston, Texas circa Hurricane Harvey. As the largest city on the Gulf Coast of the US ("Houston, Texas Population 2018" 2018), Houston had a large hurricane-affected population that, based on our analysis of Gulf-based city Tweeting behavior, also has a substantial number of affected Twitter-users.

3.3 Methods

3.3.1 Twitter Data

All of the geolocated Twitter data for the greater metropolitan area of Houston for seven weeks prior to and one week following Hurricane Harvey's landfall were streamed through the Twitter API (Wang and Taylor 2016b). Hurricane Harvey made its first landfall in Houston on August 25th in the evening; the hurricane then pivoted and returned on August 27th to deposit torrential, record-breaking rains. For our analysis, we defined our perturbed state—the period of time during which non-normal behavior would be expected—as one day prior to the first landfall through the week following landfall (August 24th-September 1st). To identify non-normal behavior, we needed to select a steady state to act as our baseline for “normal” behavior. We defined this steady state as the five-week period from July 11th and August 16th, following prior research describing the time period length necessary to generate a sufficiently stable analysis (Toepke 2018a); a longer period would increase the influence of both seasonality and population flux. The steady state behavior has a left-leaning log normal distribution, matching prior findings, and is explored more in **Figure 8**. We also allowed for a transitionary state, during which the hurricane would broadly impact Twitter posting behavior through anticipation of harm but not through actual hurricane damages or events. This state is defined as the period from the day Harvey was identified as a tropical storm through the day before our perturbed state begins (August 17th-August 23rd).

With respect to important dates, it should be noted that Houston experienced the most infrastructural damage and flooding on August 27th and not when Hurricane Harvey first made landfall. As such, many of the following analyses focus on behaviors identified on August 27th.

The sets of steady state and perturbed state Tweets were temporally aggregated by day, transformed into individual points through ArcGIS, and plotted using their latitude and longitude attribute information in ArcGIS. The Zip Code Tabulation Areas (ZCTAs) and 2010 census tract shapefiles were downloaded from the Harris County GIS data portal (Harris County, 2019).

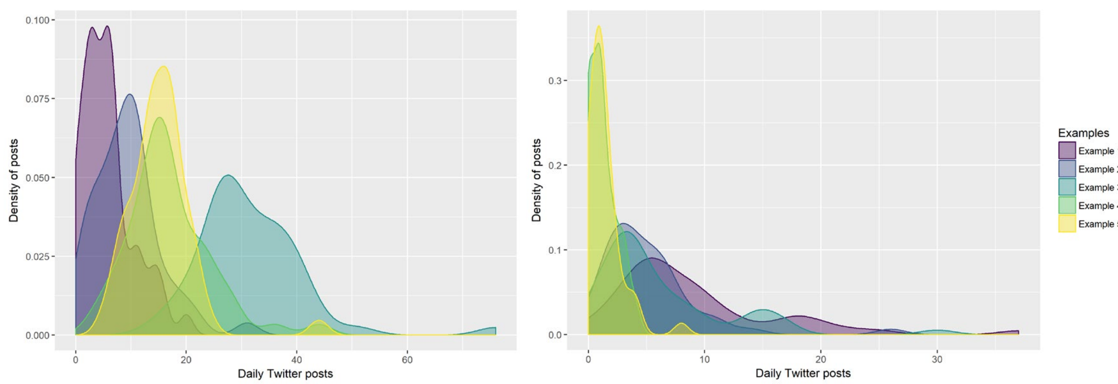

Figure 8. The steady state distributions of 5 random areas within the 50 km^2 areas spatial net (left) and the 1 km^2 areas spatial net (right).

3.3.2 Population Data

The census data and census tracts are not at a sufficiently fine resolution to enable understanding of the nuances of neighborhood-scale behavior during a crisis. The tracts further from the city center can be as large as 1500 km^2, so we need to find a method of increasing the resolution of the population data. The geographic information science (GIS) field has historically utilized National Land Cover Database (NLCD) data to increase the granularity of census data with substantial accuracy (Reibel and Agrawal 2007). The NLCD contains a raster file with 30 meters (m) by 30 m cells that have been classified, through satellite imagery, as one of 16 classes. The classification includes four classes of developed land: open space, low intensity, medium intensity, and high intensity. These classifications are determined using the amount of water-permeable and water-impermeable land; thus, the “open space” designation does not necessarily describe areas with no people, but rather areas with a relatively smaller (<20%) amount of concrete-covered land, like suburbs. The raster cells from the 2011 dataset that were classified as “developed” were extracted and clipped to the greater metropolitan Houston Area. Using ArcGIS' Raster to Point function, each of the raster cells were transformed into points located at the center of each cell and spatially joined by count into the census tracts for Houston. Using the counts of each type of NLCD class and the population record for each census tract, a multiple linear regression analysis was performed to determine the expected contribution of each type of land type to the tract’s population. The model is presented as Eq. 2.

$$Pop_t = \beta_0 + \beta_1 OpenSpace_t + \beta_2 LowInt_t + \beta_3 MedInt_t + \beta_4 HighInt_t + \epsilon \tag{2}$$

In Eq. 2, Pop_t is the population for a given census tract t; $OpenArea_t$ is the number of NLCD cells described as “Open Space (Developed)” within the census tract; $LowInt_t$ is the number of cells described as “Low Intensity (Developed)” within the tract; $MedInt_t$ is the number of cells described as “Medium Intensity (Developed)”; and $HighInt_t$ is the number described as “High Intensity (Developed)”. The results of the regression analysis are presented in **Table 4.** The model has an adjusted R-squared of 0.8317 and a model p-value of <0.001.

Table 4. Multiple linear regression results for the NLCD land class types and the census data.

NLCD Class	Model Coefficient	STD Error	p-value	Re-scaled Coefficient
Open area	0.30	0.03	<0.001***	0.23
Low intensity	0.11	0.07	0.123	0.19
Medium intensity	4.50	0.10	<0.001***	0.58
High intensity	-2.34	0.12	<0.001***	0.01

The model coefficients were used to determine the weighted averages of each land type within each census tract. As shown, the areas of very intense development have a substantial and significant negative contribution to the residential population of the Houston census tracts. Bian and Wilmot encountered similar results in New Orleans, Louisiana in a study using the same technique to study disadvantaged populations impacted by Hurricane Katrina (Bian and Wilmot 2017). Following ground proofing techniques, they assigned a positive but very small assigned coefficient to the highly developed areas of the city. Following their example for the purposes of the population disaggregation, we determined the ratio of each model coefficient to the coefficient range (-2.34 to 4.50) and used this scaled ratio as the re-scaled coefficient. This process ensured a match between the census data population and the population assigned within each tract such that the disaggregation would be at least as accurate. It also preserved the ratio of the magnitude of impact between categories.

The assigned coefficients were employed as weighted averages in Eq. 3 to determine a population quantity to assign to each of the NLCD point shapefiles.

$$Pop_{cell_i} = \frac{\frac{WA_i}{\Sigma WA_{i \to l}} * Pop_{total_{tract}}}{Tract_{count_i}} \tag{3}$$

In Eq. 3, Pop_{cell_i} is the population represented by a single 30mx30m NLCD raster cell point of NLCD type i within a specific census tract; the NLCD types of “Open area”, “Low intensity”, “Medium intensity”, and “High intensity” are represented as $i \to l$; WA_i is the weighted average for the specific land type; $WA_{i \to l}$ is the sum of the weighted averages for each land type; $Pop_{total_{tract}}$ is the total population within the census tract in which the point is located; and $Tract_{count_i}$ is the total number of points of type i within the specific census tract. The resulting dataset was a grid of point files spaced 30 m apart that could be aggregated into equal-area, uniform areas with less structural bias than census tracts. Although the accuracy of this method cannot be accounted for at a 30mx30m scale, the regression results and the rescaling of the coefficients lend the assurance of reasonable accuracy for larger-scale aggregation.

3.4 Spatial Nets

A common method of aggregation, particularly when social factors are considered, is to use census tracts or ZCTAs (Grubesic and Matisziw 2006). Although this gives arguably the most accurate nighttime population of the aggregated areas, their boundaries can be of widely varying shapes and sizes, and their jurisdictions are strongly influenced by the boundaries of socially-biased fragmentations. These social factors and can contribute data bias, and the size and shape differences can contribute to the Modifiable Area Unit Problem (MAUP) (Grubesic and Matisziw 2006; Jelinski and Wu 1996). As such, we chose to create our own spatial nets of varying scales and to include the ZCTA boundaries for comparison.

With the social media data temporally aggregated and mapped and the population data spatially disaggregated, we designed twelve spatial nets to catch the population data and the Twitter stream for each day. These spatial nets are composed of a series of interlocking shapefiles that cover the greater Houston area. One of these nets consists of the Houston ZCTA zones; the others were composed of uniformly shaped, tiled hexagons. As many of the problems identified with using manmade boundaries for spatial aggregation are related to their varying sizes and shapes, we wanted to develop a spatial net design that could be deployed across a large area, regardless of country of interest, and could be scaled according to the intended research design. This has the advantage over ZCTAs, which vary in size and shape. The ZCTAs within Houston alone, for instance, range in size from 0.16 square kilometers (km^2) to 677.20 km^2.

We generated the spatial nets of equally sized and shaped hexagonal polygons through ArcGIS' Generate Tessellation function. Hexagons are better suited for tiling large geospatial areas because they reduce edge effects that can be exacerbated by intersecting rectangles and are more scalable on a curved surface like the globe (Carr et al. 1992; Polisciuc et al. 2016). The twelve hexagonal nets consist of hexagons that have square areas of, respectively, 0.25 km^2, 0.5 km^2, 0.75km^2, 1 km^2, 2 km^2, 5 km^2, 10 km^2, 15 km^2, 20 km^2, 35 km^2, 50 km^2, and 80 km^2. These areas were chosen following the guidelines listed in the Spatial Association of Scalable Hexagons described by Potter et al., which suggests choosing sizes that closely mimic the behavior that is being studied or the sampling size, or the spatial dependence of the data. As such, these sizes mimic the range of sizes of Houston census tracts and the spatial spread of the Twitter data. With respect to the upper limit of 80 km^2, we determined from the stated interests of emergency management personnel that information on areas larger than that scale provided very little actionable or useful information in terms of aid distribution or the presence of local disasters. A comparison between the 1km^2 and 80km^2 nets is shown in **Figure 9** for a scalar reference.

We summed the number of Twitter posts for each day and the population values of each of the NLCD points within each polygon of each net. Following the methods listed Kryvasheyeu et al. (2016), polygons that did not contain at least one Tweet per day during either the steady state or perturbed state were removed. For validation purposes with respect to the hurricane damage, we additionally plotted the Federal Emergency Management Agency building level damage assessments (FEMA damage assessments) collected in the days following Hurricane Harvey. These assessments are geolocated and the buildings they reference are classified as "Affected", "Minorly Affected", "Majorly Affected", and "Destroyed". We converted this ordinal scale of damage into a numerical ordinal scale, for which "Affected" is classified as a "1" and "Destroyed" is classified as a "4". For each polygon of each net, we additionally calculated the total number of damage assessments performed and the maximum and average assigned damage value of the polygons.

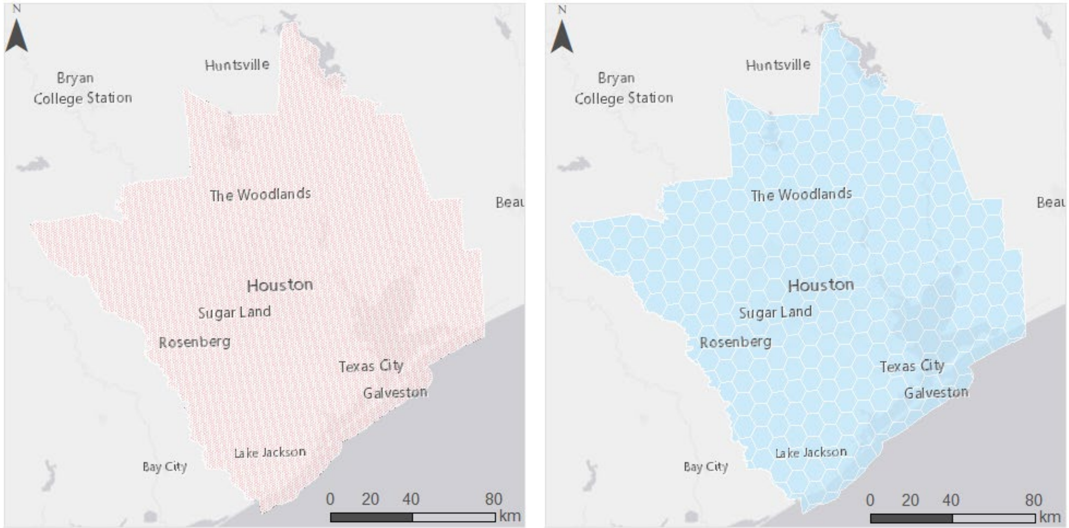

Figure 9. (Left) Spatial net consisting of 1 km^2 hexagons across Houston. (Right) Spatial net consisting of 80 km^2 hexagons across Houston.

3.4.1 Analytical Methods

H1. The distribution of changes in social media behavior and the identification of behavioral clusters is statistically different at smaller scales.

In order to identify the scalar interval of aggregation at which the distribution of Tweets per person changes significantly, we created empirical cumulative density functions (CDFs) of the Twitter activity per person for each net on a daily basis. These intra-net CDFs consisted of all of the Twitter representation values identified on a single day in all of the polygons of the net. We produced thirteen CDFs for five days of the steady state and for each day of the perturbed state. We then used the Kolmogorov-Smirnov II test on each pair of CDFs to test the likelihood that the CDFs were produced from the same parent distribution (Massey 1951).

H2. The identification of crisis-induced, extremely high or extremely low amounts of Twitter activity is scale-dependent.

In order to understand the distribution of perturbed state Twitter posting counts that were either much higher or much lower than the "normal" behavior observed in the steady state, we used cumulative distribution functions (CDFs) to compare perturbed state Twitter activity to the steady state Twitter activity. We used the steady state post counts for each of the spatial nets to generate a series of CDFs. These CDFs represented the distribution of Twitter activity for a single area across each day of the steady state period. For instance, for a given area **A** within the 10 km^2 net, we created a CDF from all of area **A**'s steady state Twitter activity counts by day. We then took the perturbed state Twitter activity on a given day, such as August 27th, and used the generated steady state CDF of activity to determine what percentage of steady state days had produced less than the number of Twitter posts produced on August 27th in area **A**. A result of 0.90 would indicate that the perturbed state activity on August 27th was higher than the activity produced on 90% of the days in the steady state, and a result of 0.10 would indicate that the perturbed state activity was only higher than 10% of days in the steady state.

We used each CDF to assess, for each area and each day of the perturbed state, the cumulative likelihood of observing a certain number of Tweets in that area on that day. We categorized this likelihood as being normal, non-normal, or extreme. Although, as stated, we are most interested in extreme values, we included an analysis of non-normal to provide a reference for the impact of how the threshold of "extreme" amounts of activity is defined. Using empirical rule values, non-normal social media behavior was defined as being less than 16% of steady state values or greater than 84% of steady state values. Extreme social media behavior was defined as being less than 5% of steady state values or greater than 95% of steady state values. To identify the effect of scale on observing extreme (and non-normal) values and so understand the prevalence and significance of activity bursts, clustering, or drop-offs, we took the distribution of the likelihood of observing the perturbed state Twitter activity and analyzed the distributions of those probabilities across nets and days of the perturbed state.

H3. The strength of the correlation between Twitter activity and social media is scale-dependent.

Finally, to assess how scalar aggregation affects the previously identified correlation between non-normal Twitter activity and hurricane damage, we applied the statistical test for Kendall's rank coefficient to the Twitter activity per person within each area and the average FEMA building level assessment designation recorded within each area ("Kendall Rank Correlation Coefficient" 2008). These values were determined on a daily basis and compared across spatial nets.

3.5 Results

H1. The distribution of changes in social media behavior and the identification of behavioral clusters is statistically different at smaller scales.

The comparisons of the Kolmogorov-Smirnov II test results are depicted in **Figure 10a-d**. The null hypothesis for the test is that the two populations are drawn from the same population. With larger p-values (distribution pairs with p-values greater than 0.05 are displayed as gray squares), the less certain we are that we can reject the null hypothesis The test statistic for the Kolmogorov-Smirnov II test is the D-value, which is representative of the greatest distance between the two tested distributions. The D statistic is larger when the statistical difference between the two CDFs is greater, and is represented by a more teal than purple hue.

The top two figures show the results of two days of the steady state, August 7th and August 8th, and the bottom two figures show the results of two days of the perturbed state, August 26th and August 27th. It should be noted that the CDF and statistical test results for the census tracts are on the outermost edge of the figures, and those results, interestingly, are not equivalent to either the very small nets or the very large ones. Across each set of tests, the larger nets' distributions are more similar to each other, and nets that are more similar to each other in size also have more similar distributions. Apart from those trends, both of the steady state graphs show the significant decrease in the certainty that the spatial nets' CDFs are from different distributions between 5 km^2 and 2 km^2. This pattern of 2-5 km^2 spatial trends breaks during the perturbed state, in which small-scale areas begin to behave more similarly to each other. This opened window in the smaller spatial scales exists from August 26th through September 1st, although it diminishes in size beyond August 29th. It is largest on the day of maximum damage from rainfall, August 27th.

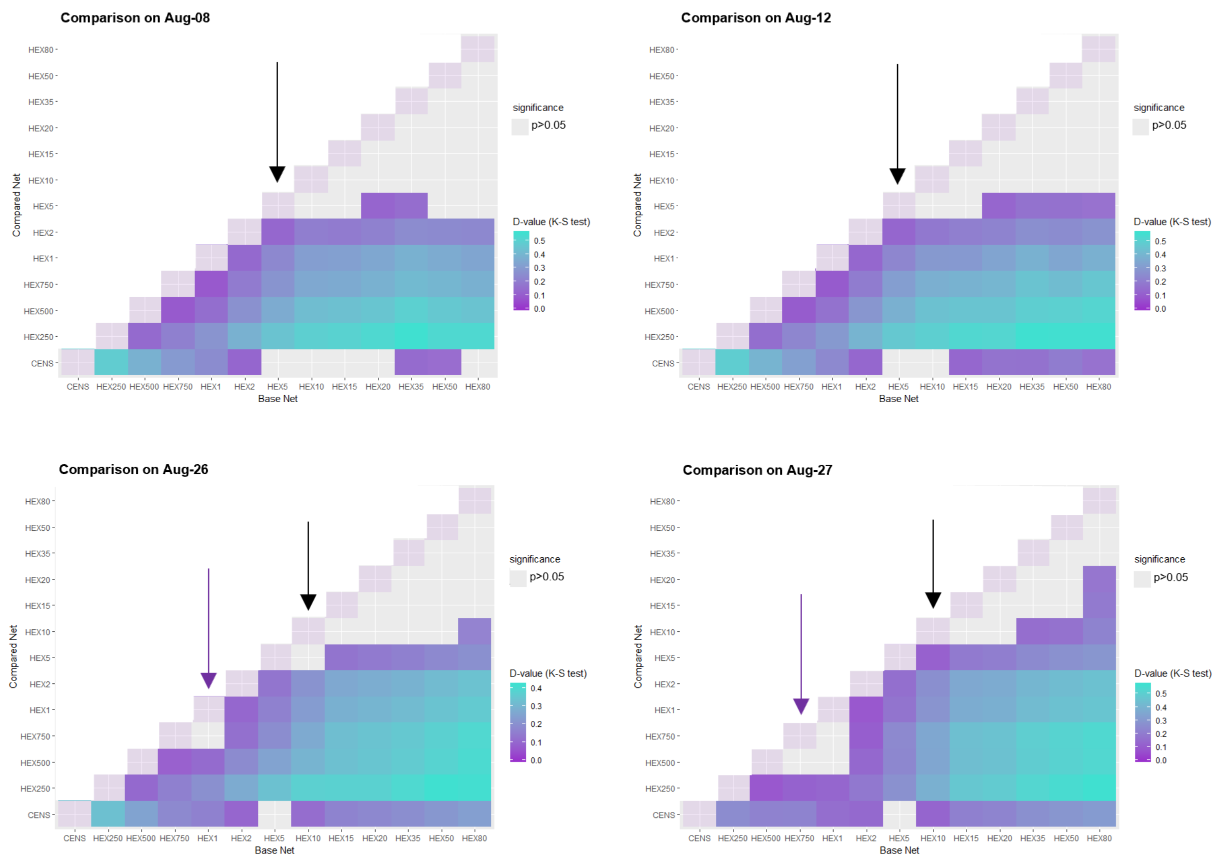

Figure 10a-d. Comparison of the likelihood that the spatial nets have different daily distributions using the Kolmogorov-Smirnov II test. The census tracts net is represented in the furthest left column and bottom row. The top two figures were created from steady state values (August 8^{th} and August 12^{th}), and the bottom two figures were created from perturbed state values (August 26^{th} and August 27^{th}). The black arrows indicate the first pair of nets, by increasing scale, that are statistically distinct from the next smallest net. The purple arrows indicate pockets of statistically indistinct distributions at the 0.5 - 1 km^2 scale that appear in the perturbed state.

H2. The identification of crisis-induced, extremely high or extremely low amounts of Twitter activity is scale-dependent.

In order to understand the distribution of perturbed state values that were either much higher or much lower than the "normal" social media behavior observed in the steady state, we used CDFs to compare perturbed state values to the steady state values. For instance, for a given area A, within the 10 km^2 net, we created a CDF from all of that area's steady state values. We then took the perturbed state Twitter activity on a given day and used the CDF to determine what percentage of steady state values were lower than that perturbed state value. A value of 0.90 would indicate that the perturbed state value was higher than 90% of the steady state values. We then looked at the distribution of those probabilities to identify the effect on scale on the prevalence of extreme or non-normal values. First, we looked at the distribution of these probabilities graphically. The densities of the intra-net daily CDFs for the spatial nets for 0.25 km^2, 1 km^2, 5 km^2, and 35 km^2 are portrayed in **Figure 11a-d**. **Figure 11a** and **b** show bimodal distributions, with increased numbers of areas displaying extremely low and high amounts social media engagement. **Figure 11c**, representing the 5 km^2 spatial net, has a less ordered bimodal distribution, with less consistency across days, and **Figure 11d** (35 km^2) does not exhibit a bimodal distribution, and is remarkably disorderly. The peak density values also decrease with increasing spatial net size.

Following a qualitative assessment of distribution, we quantitatively assessed the percentage of values on the day of maximum rainfall and damage, August 27th, that exhibited non-normal or extreme social media behavior. Non-normal behavior was defined as being one standard deviation from the mean, i.e, less than 16% of steady state values or greater than 84% of steady state values. Extreme social media behavior was defined as being less than 2.5% of steady state values or greater than 97.5% of steady state values. We additionally included the percentage of areas that exhibited normal Twitter activity, within the central 68% of observed Twitter activity, on both figures for comparison.

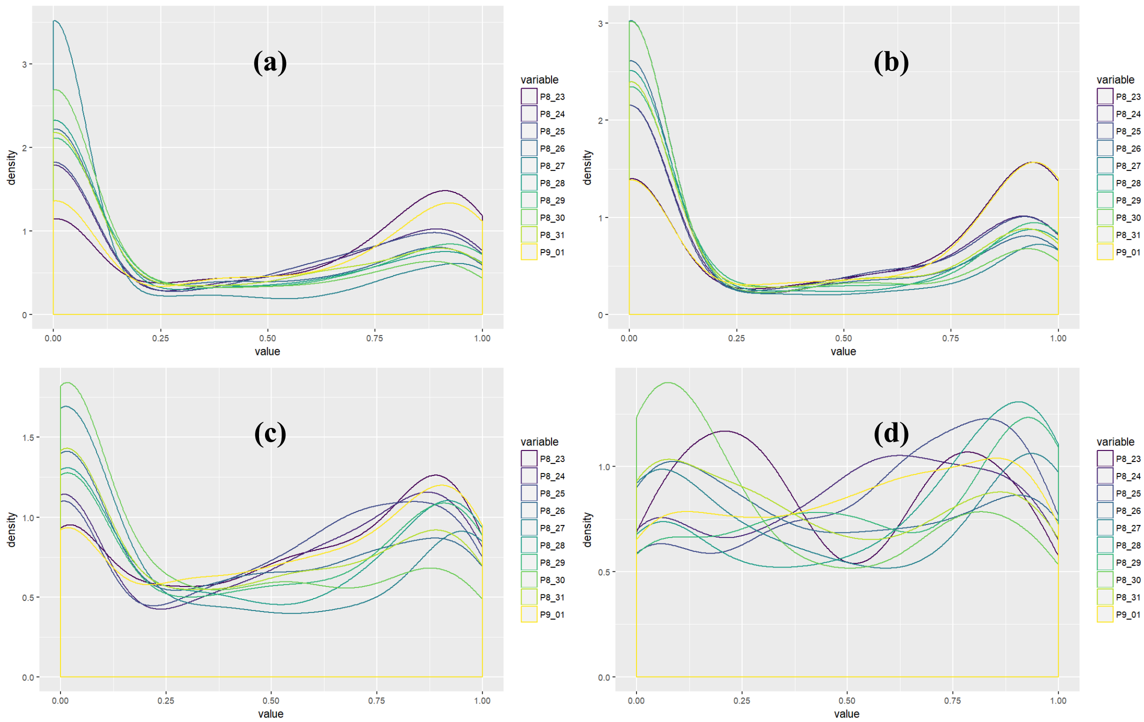

Figure 11a-d. Comparison of the densities of extreme and non-normal social media behavior for four spatial nets (0.25 km^2, 1 km^2, 5 km^2, and 35 km^2). The colors represent different days of the perturbed state, and peaks closer to 0 or 1.0 indicate values that are either lower or higher than most of the values observed in the region's steady state.

The percentage of areas identified with each of those kinds of social media interactions compared to the size of the spatial nets are compared in **Figure 12** and **Figure 13**. To define the relationships between each set of variables, we used the method of maximum likelihood to estimate a scaling exponent for a power law relationsship. A power law relationship is defined as $f(x) = \alpha x^{\beta}$; with this data, x represents the geographic scale, and $f(x)$ indicates the predicted variable (i.e., the percent of areas with extremely high or low numbers of Tweets, the correlation coefficient between Tweets and FEMA damage ratings, or the percentage of geographic area excluded from analysis), and α and β are constants (Stumpf and Porter 2012). The minimum x value (geographic scale) was determined by minimizing the Kolmogorov-Smirnov test statistic, and the goodness of fit of the estimated parameters for these relationships were assessed using the Kolmogorov-Smirnov test (Clauset et al. 2007). Although there are too few data points to more positively confirm the presence of a power law relationship, the Kolmogorov-Smirnov test statistics show that the data distributions for each of the tested relationships could be derived from a power law distribution. Vuong's test was used to compare the relative distance between the sample distributions, the estimated power law distributions, and log-normal distributions estimated from the same data (Vuong 1989). The estimated power law scaling exponents, the minimum x values, the Kolmogorov-Smirnov test statistics, the Kolmogorov-Smirnov test results, and the Vuong test results are presented in **Table 5**.

Of most importance for applications, the relationship between increasing geographic scale and the identification of extremely low social media interaction is negative, while the relationship between scale and the identification of extremely high social media interaction is positive. The minimum value for which this relationship holds true has not been identified. The scaling constants, α, vary between 0.1 and 1.0, but the magnitudes of the power β is approximately the same for both of the equations for the extreme values. Additionally, the inclusion of non-normal activity analysis was provided as a reference for the impact of the threshold at which the amount of activity could be interpreted as "extreme". We see the impact of increasing the boundary for the classification from outside the central 68% to outside of the central 95% in the identification of non-normal and extreme perturbed state activity for both sets of behavior trends.

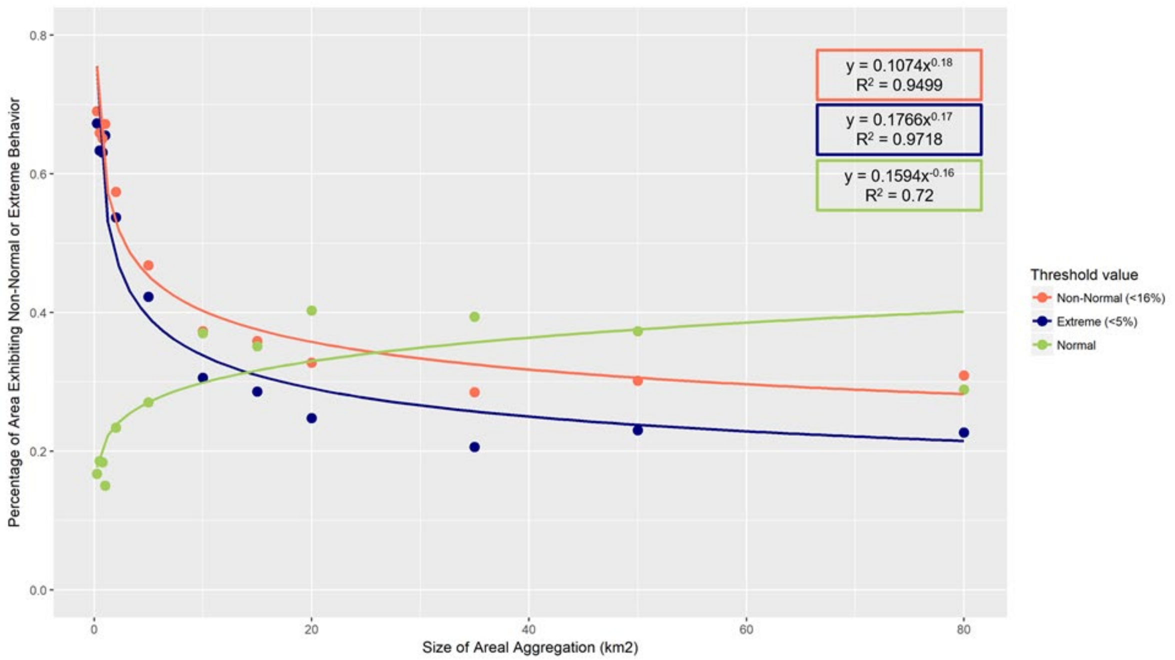

Figure 12. Comparison of the percentages of areas exhibiting non-normal or extremely low Twitter activity behavior between spatial nets, fit to power law distributions.

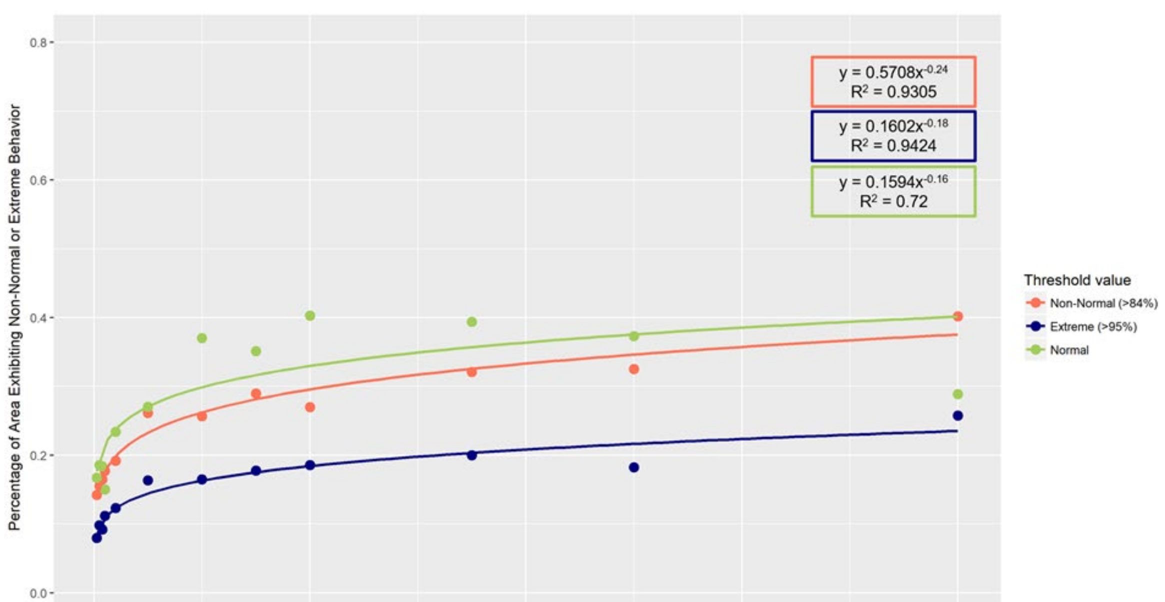

Figure 13. Comparison of the percentages of areas exhibiting non-normal or extremely high Twitter activity behavior between spatial nets, fit to power law distributions.

Decreasing the threshold at which an observed behavior is classified as noteworthy obviously increases the number of noteworthy observations; however, this effect is slightly larger for the extremely high values, and the effect is more profound at higher scales (greater than 40 km^2).

At a gestalt level, the difference between the percentage of areas identified as exhibiting extreme at the smaller scales is much larger (approximately 80% for scales less than 0.5 km^2) than the percentage identified at the larger scales (approximately 53%). This relationship is represented in **Figure 14**. The geographic coverage of those areas, however, is quite similar due to the increased removal of areas without Twitter activity at the smaller scale.

H3. The strength of the correlation between Twitter activity and social media is scale-dependent.

The relationship between scalar aggregation and the strength of the correlation between Twitter activity on August 27th and hurricane damage is presented in **Figure 15**.

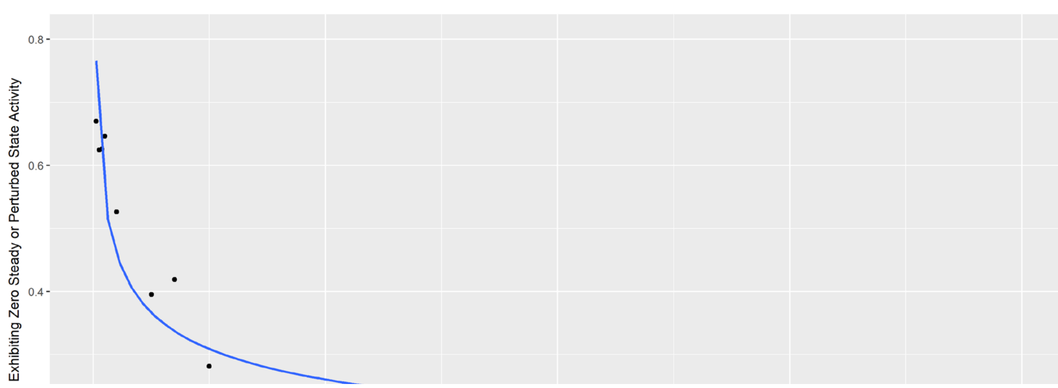

Figure 14. The relationship between the percentage of the total study area (the greater metropolitan area of Houston) excluded from the analysis on account of not having sufficient Twitter activity, as defined in Section 2.3, and the geographic scale at which the data was aggregated.

This data also shows a direct proportionality between the scale of aggregation and the strength of human social media behavior signals in the analysis. It should be noted that each value is significant except for that of the 35 km^2 net, which was excluded from the model fitting due to its lack of statistical significance ($p = 0.11$) . The correlation strength increases sharply until the 10 km^2 scalar aggregation with no apparent sacrifice of statistical significance.

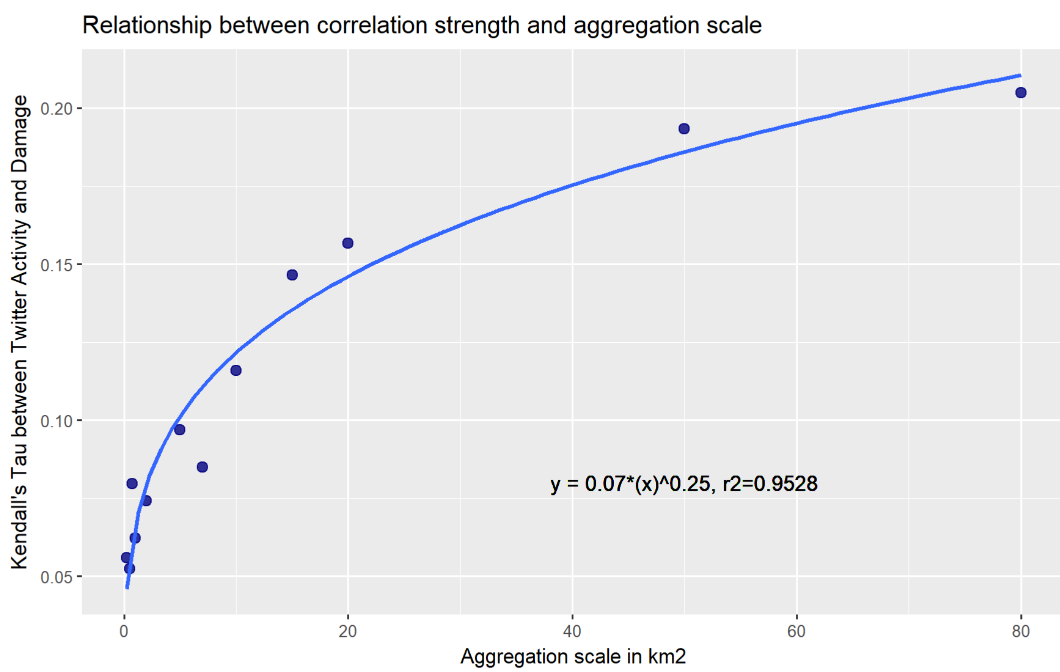

Figure 15. Kendall Rank Correlation Coefficient strength between Twitter Activity per person and the average FEMA building damage assessment rating within each area.

Table 5. Results of the maximum likelihood estimation of sample distribution fit to power laws.

Variable	β	x_{min}	Exponent	K-S test	Vuong's test
Extremely high activity	0.16	500m	-0.18	0.27*	Power law more likely
Non-normally high activity	0.57	500m	-0.24	0.26*	Power law more likely
Non-normally low activity	0.1	250m	0.18	0.16*	Log-normal more likely
Extremely low activity	0.18	250m	0.17	0.17*	Equally likely
% of area removed	0.58	250m	-0.32	0.16*	Power law more likely
Correlation with damage	0.07	500m	0.25	0.16*	Equally likely

*Indicates that the Kolmogorov-Smirnov test statistic's null hypothesis (that the sample distribution was drawn from a power law distribution) could not be rejected at $p<0.05$.

3.6 Discussion

First and foremost, the analyses presented herein show that the stories our data tell differ when read at different scales. With any scalar analysis, we would expect more variation at smaller scales across time and between areas. There is an expected tradeoff between certainty at large scales and specificity at small scales. The crisis community has recognized this; however, the effect of scale on the ability of social media to (**H1**) identify distinct clusters of geographic social media interaction changes, (**H2**) identify non-normal or extreme social media behaviors, and (**H3**) provide statistical confidence that social media behavior changes indicate danger had not been explored. Within our three analyses, we have identified the precise relationship between scale and social media signal behavior in the hopes of making the crisis informatics community more aware of how scale can influence multiple facets of the findings of social media analytics.

In addressing **H1**, the steady distribution of Twitter activity observed within spatial nets of an area greater than 10 km^2 is not statistically different from the distribution observed in another spatial net greater than 10 km^2. The inverse is also true; nets smaller than 5 km^2 are not as likely to be from the same distribution as each other. This confirms the existence of social and place-related social norms occurring at scales smaller than 5 km^2; i.e., in the broader scheme of human activity, most locations and events occupy a geographic space smaller than 5 km^2, so smaller scale analyses incorporate different slices of these different locations and events. At the 10 km^2 scale, the highs and lows of activity are averaged across more data, minimizing the impact of the grouped extremes. Even at the 0.25 km^2 scale, the analysis identifies different slices of life from the 0.5 km^2 scale.

This clear cut-off of statistical difference disappears during a crisis state, however. People within 0.5 – 2 km^2 begin behaving in more similar ways, and, during the day of maximum rainfall and damage, the difference between the distributions at most of the smaller scales decreases substantially. This confirms research showing hurricanes impacting cities differently at small scales due to small vortices and flooding susceptibility. It is likely that this sudden homogenizing of social media behavior at the 0.5 – 2 km^2 is indicative of the average effect of distinct, human-impacting hurricane phenomena. This similarity of social media behaviors at smaller scales than usual may also be indicative of reduced population mobility or infrastructure limitations.

With respect to our second hypothesis, the percentage of extreme social media behavior in a crisis state and the strength of the correlation between extreme behavior and hurricane damage are both definably dependent on geographic scale. We identify and define the effect of geographic scale on the identification of extreme social media behaviors, and we show clearly that the effect is different for different extreme behaviors. When we assessed the distribution of the likelihood of seeing each value in the perturbed state using the empirical cumulative distribution function generated in the steady state, we found that the smallest scales are likely to define the majority of active areas as either very low or very high. Few perturbed state values lie close to the average of the steady state. If the purpose of social media analysis is to identify areas with higher or lower severity, this indiscriminate binning of most values as "extreme" would not be ideal. The consistency across the perturbed state does, however, note a reliably consistent categorization of specific areas into extremes. This is the inverse of the social media behavior observed at the larger spatial scales, which has a wide variety of probabilities of the occurrence of values, and yet varies drastically from day to day of the perturbed state.

The effect of geographic scale on variability in data distributions is reinforced by the relationships between aggregation scale and the number of identified non-normal and extreme events. In our data, we have identified six possible power law distributions, although three of them are more likely to truly be drawn from a power law distribution than the others. Power law distributions have attracted a large amount of attention in almost every field, ranging from microbiology to economics. They have been suggested as being present in nearly every natural system, although the statistical confirmation of the reality of each claim has been questioned (Stumpf and Porter 2012). The key feature of a power law relationship is that it is independent of scale: the relationship between the two variables is constant (although that relationship must often be restricted by a minimum value of the independent variable). This scale-invariance can be seen to indicate an intrinsic characteristic of the system.

In the realm of social behavior and social networks, power law relationships have been shown to develop due to growth and preferential attachment (Barabási and Albert 1999). The prevalence of those two features in human networks has contributed to the identification of many social power law relationships, such as the famous "rich-get-richer" phenomenon noted by Pareto, who was one of the pioneers of power law identification (Newman 2005). The flux of people into urban centers can easily account for growth, and preferential attachment can be seen in how social media-using demographics more commonly flock to urban centers with similar types of people (Shelton et al. 2015).

Recent work has identified additional power law relationships between the population size of city centers and the number of different types of Tweets generated within (Fan et al. 2020). Those results, which stem from the number of people in specific cities, and our results, which stem from the number of people within geographic scales, seem to indicate that growth and preferential attachment are present in the relative spatial clustering of those who want and are *able* to use social media more during a disaster. Because of these qualities, despite continuous growth in urban populations and social media users, the network of users specifically using social media during a disaster is organizing itself into a scale-free clustered network. This phenomenon may affect the relative ability of areas with greater numbers of social media users to receive more resources during a disaster purely due to the concentration of their voices.

It should be noted, however, that a scale-free stationary state is difficult to prove (Forman 2007). Criticisms of abundant labelling of power law dynamics require statistical tests that our data cannot satisfy, i.e., independent and dependent variables that range more than two orders of magnitude each (Stumpf and Porter 2012). Additionally, many statistical tests for the goodness of fit of power law distributions for binned data additionally require a longer tail than our data provides (Virkar and Clauset 2014). With the limited discrete data obtained in this study, we are unable to determine with certainty the existence of power law relationships. We have followed the statistical suggestions put forward by fervent critics of the search for power laws (Clauset et al. 2007; Forman 2007) as well as we could. For two of our defined relationships, we cannot show through Vuong's test that a power law distribution better describes our data than a log-normal or exponential distribution. Because of this uncertainty, the results of this study should be limited to the highest xmin determined through minimization of the Kolmogorov-Smirnov test statistic (500m) and the maximum scale we utilized (80km).

We can at least say with certainty that the number of identified extremely low events decreases exponentially with increasing scale, and the number of identified extremely high events increases exponentially with increasing scale. This increased identification of non-normal social media interaction at increased scales suggests the need to apply more stringent thresholds for activity marked as abnormally or extremely high at larger spatial scales. That said, the many decreased values identified at the smaller scale call for more stringent methods of investigation into these areas that are suddenly silent.

In terms of sudden silence, previous research has identified that drop-offs in Twitter activity are also correlated with damage and theorized that those drop-offs are caused by social vulnerabilities more than social media behavioral choices (Samuels et al. 2018b). Increased scales minimize the potential for a social media analysis to identify these drop-offs as extreme events, a factor that needs to be considered and addressed in social media applications.

In applying these to future analyses, we also show the influence of the MAUP on the use of ZCTAs in crisis informatics. ZCTAs vary widely in size and shape; the ones within Houston, for instance, range in size from 0.16 km^2 to 677.20 km^2. This variance in size and socially constrained boundaries have been substantially critiqued in the field of critical GIS (Jelinski and Wu 1996; Saib et al. 2014). Across all analyses, the distributions appear closest to the values for the 5 km^2 net. As the average size of the census tracts for the area is approximately 7 km^2, this suggests that the potential spatial biases of census tracts in terms of Twitter representation may be more directly related to the tracts' geographic size and less their socially-constructed boundaries. The variances in the sizes of the ZCTAs is an additional variable that, as we have shown, has a significant effect on the analytical results of an analysis.

As for the third hypothesis, concerning the relationship between these extreme values and hurricane damage, there have been multiple remarks in the literature regarding an expected relationship. Shelton et al. (2014) identified a discrepancy in correlation values at varying scales, noting the apparent necessity of including scale as a factor in any analysis comparing Twitter activity and hurricane damage. County-scale and state-scale correlations have found to be moderately strong (Guan and Chen 2014; Kryvasheyeu et al. 2016; Shelton et al. 2014); however, each author notes the influence of scale on their analyses. Guan and Chen hypothesize that "moving upward on the scale is likely associated with a larger amount of disruptions at a higher level of severity", which would lead to a stronger, more significant "disaster" signal. Within our analysis, we are able to show the likelihood of a power law relationship between increasing analytical scale and the strength of the correlation between damage and Twitter activity. This relationship shows that we can be more confident in social media activity behaviors indicating a local hazard when we look at larger spatial scales.

In our data, we have identified six potential power law distributions. Power law distributions have attracted a large amount of attention in almost every field, ranging from microbiology to economics. They are nearly ubiquitous in natural systems, although the statistical confirmation of the reality of each claim has been questioned (Stumpf and Porter 2012). It is no small irony to the authors that their identification of the scale-dependence of social media analyses culminates in a distribution defined as scale-independent. We would like to note, then, that the relationship between scale of aggregation and these Twitter behaviors is what is scale-independent; the relationship itself is direct and significant.

The significance of the power law relationship itself identified between these variables is less certain. Criticisms of abundant labelling of power law dynamics require statistical tests that our data cannot satisfy, i.e., independent and dependent variables that range more than two orders of magnitude each (Stumpf and Porter 2012). Statistical tests for the goodness of fit of power law distributions for binned data additionally require a longer tail than our data provides (Virkar and Clauset 2014). Ultimately, whether these relationships are power law or logarithmically distributed, we see an exponential or as-good-as increase in the correlation between Twitter activity and hurricane damage. The certainty involved in whether extreme social media behaviors function as good indicators of hurricane damage is scale-dependent, showing once again that the tradeoff for geographic specificity is certainty of the identification of hurricane damage, and analyses performed at the census tract or county level need to incorporate the analysis scale into their explanations of their findings.

3.6.1 Limitations and Future Work

NLCD data. Although one of the first uses of NLCD data was to estimate populations distributions on a fine scale, and despite its storied history in population estimation, it is still a fallible metric. Despite the good fit of our initial regression model, the negative coefficients produced required additional model tweaking that undoubtedly caused further data bias. Additionally, census data is limited to the “nighttime” population, i.e., where people sleep (home). Based on the Tweet text, we find that most people stayed home during the worst of the hurricane, and thus this nighttime population may double as hurricane-time population. We were incapable of incorporating evacuation dynamics for the hurricane, but evacuation orders were issued too little and too late for Houston. Evacuation likely had less impact on population dynamics for Houston than for areas more often affected by hurricanes.

Hurricane-specific Tweeting. Many studies in the field filter for Tweets that are directly related to the hurricane through text analysis. This limits the application of steady state versus perturbed state analysis, as no one was Tweeting about Hurricane Harvey before it formed in the Gulf. We additionally wanted to incorporate areas Tweeting in a “business-as-usual” fashion during the hurricane. As such, using hurricane/disaster-specific Tweets was not possible, and our analysis undoubtedly incorporated some Twitter bots (Yang et al. 2019). We manually filtered some of the bots based on keywords (i.e., “jobs”, “FloodWatch”) determined through manual application of the OSoMe tool

BotOrNot (Indiana University 2018). Furthermore, as the bots are likely unaffected by the hurricane, their activity changes would be normalized to zero, thus minimizing their effect on our analyses. Further research is necessary into the influence of Twitter bots on the Twitter distributions and analyses specific to disasters.

Area exclusion. We excluded from analysis all areas that did not have a single Tweet across the steady state period or a single Tweet across the perturbed state. However, this resulted in a substantial amount of geographic coverage reduction in the smaller scales. We identified a logarithmic increase in excluded area with decreasing scale. For example, 66% of Houston was excluded in the 0.25 km2 net, 53% was excluded from the 1 km2 net, and only 20% was excluded from the 20 km2 net. Further research is necessary to investigate the effects of Twitter activity thresholds, population thresholds, and the effect of including or excluding areas and populations with a very lower Twitter representation. This information will help inform how applicable Twitter data can be to demographically-different neighborhood distinctions in social media behavior and will minimize the inclusion of non-participating, unrepresented populations.

Lastly, no two disasters are the same, either in terms of the damage caused or in terms of the affected society. The generalizability of the potential power law relationships and the distributions of Twitter activity to other cities and other disasters should be investigated.

3.7 Conclusion

There will never be an answer to, “What is the best scale at which to perform crisis informatics analyses?” just as there is no answer for, “At what scale are societies affected by disasters?" We hope, however, to have provided a roadmap for crisis informatics researchers using social media to better understand how the chosen scale of their analysis will affect their results. Many of the potential power law relationships identified in this paper indicate an exponential tradeoff between the geographic specificity of smaller scales and the statistical certainty that an identified social media behavior represents an endangered population. Crises within a disaster context happen to individuals and communities; it is important to work towards using social media data to improve our ability to correctly assess the severity and magnitude of an identified emergency. Both of those factors depend on geographic scale relative to the surrounding areas and that area's own history. How we ascertain and contextualize our data, then, is heavily influenced by how we structure our analysis, and we need to be wary of what dependencies might be tipping the scales

3.8 Acknowledgements

This material is based upon work supported by the National Science Foundation under Grant No. 1760645. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

CHAPTER 4. DEEPENING THE DIVIDE: CRISES DISPROPORTIONATELY SILENCE VULNERABLE POPULATIONS ON SOCIAL MEDIA[2]

4.1 Abstract

In the past decade, crisis informatics has sought to produce and use actionable information from social media data. Although substantial progress has been made in discerning how the data can be used, there is a lack of research in identifying possible inequities in that use. Previous research has shown that vulnerable populations use social media less in a disaster; however, the extent to which this social media usage disparity is predictable and the magnitude of that disparity have not been explored. This paper compares the covariance in Twitter activity and social vulnerability factors during a steady state period pre-hurricane and the perturbed state period following Hurricane Harvey's landfall. These models show that sociodemographic vulnerability factors are better in predicting Twitter activity during a crisis than infrastructural damage, that sociodemographic factors negatively influence Twitter activity, and that this phenomenon is strengthened by a crisis. The crisis-specific negative covariance indicates the need for increased consideration of vulnerability factors in social media data-driven management of urban resilience and resource distribution.

[2] This chapter has been accepted for publication by the ASCE Journal of Management in Engineering with Dr. John E. Taylor as the co-author. The citation for the journal article is as follows: Samuels, R. and Taylor, J.E. (2020). "Deepening the Divide: Crises Disproportionately Silence Vulnerable Populations on Social Media." *Journal of Management in Engineering*. https://doi.org/10.1061/(ASCE)ME.1943-5479.0000848

4.2 Background

As new and varied forms of information become available to researchers during crises, there has been a substantial push towards finding ways of applying that information to emergency responder priorities on the ground and at higher levels of decision-making. More people than ever are living in areas susceptible to catastrophic disasters because of urban sprawl (Allen 2006) and increased extreme weather patterns from climate change (Adachi et al. 2017). As such, our ability to effectively and accurately utilize all forms of available information will be critical to reducing loss of human life and increasing the resilience of our cities. While worsening extreme events are becoming more of a certainty than a possibility (Hauer et al. 2016), the extent of the impact on humans and society can be mitigated through improved resource planning and resource agility. These can be improved through increased real-time information on human location, activity, and *in situ* capabilities (Roshan et al. 2016). Ultimately, more efficient distribution of our resources will depend on what is known about the people caught in the path of these extreme events.

One source of data on human social media behavior and the on-the-ground information is the data generated through human interaction with communication networks. Social media has been found to offer new angles of study for individuals' thoughts and sentiments across broad ranges of applications, from the construction industry to crisis resilience (Reuter and Kaufhold 2018; Tang et al. 2017). Data sources such as Twitter (Spence et al. 2015), FourSquare (Aubrecht et al. 2017), and cellular data (Jennex 2012) are particularly useful as they each can have unique user identifiers, a location attribute, and a topical attribute, such as the text of a Tweet or the type of store someone has visited. The incorporation of user-volunteered information has been useful for tracking individuals' mobility and the influence of a disaster on that mobility (Wang et al., 2017), the change of individuals' sentiment in response to different disaster impact levels (Wang and Taylor, 2018), and to identify infrastructure service disruptions using social media data mid-disaster (Fan and Mostafavi 2019). Research using spatiotemporal aggregation to compare two spatial datasets has shown that bursts of social media behavior and disaster-related posts can indicate areas of relatively higher hurricane damage (Kryvasheyeu et al. 2016) and the location of flooding (de Albuquerque et al. 2015). Looking forward, user-volunteered information has been proposed for use in digital twin city frameworks for assessing infrastructural vulnerabilities, thus improving disaster resilience and preparedness (Xu et al. 2016), and for improving the situational awareness of emergency responders using the digital twin through integrated text, image, and geopositioning analysis (Fan et al. 2020).

As compelling as these findings are, big data research has often been critiqued for overlooking human variability and for mistaking big data for complete data (Blumenstock 2018; Gandomi and Haider 2015). These two fallacies can also be found intertwined in some aspects of existing crisis informatics, as one of the critical dilemmas with humans-as-sensors analyses is that humans are not reliable sensors. Humans do not transmit consistent, coordinated, or comparable information through public data channels that can be continuously accessed by connected emergency responders or data analysts. The rush to utilize information produced by humans-as-sensors in disasters has neglected to incorporate the diversity of human response and capabilities, impairing proper management and stewardship of that information.

Data bias in analyses involving Twitter data has been found in both the sample population (available Tweets) and the algorithms that have been constructed to parse the Twitter information (Johnson et al. 2017). The types of Tweets that are most used by researchers—those with a geotag—contain an additional set of biases based around the types of people that intentionally choose to use a geotag (Malik et al. 2015), further muddying the waters. Perhaps most concerningly, in terms of population bias, one study found that 50% of deaths from Hurricane Sandy occurred in an area with a complete lack of Twitter activity (Shelton et al. 2014). Another study focused on Hurricane Harvey found that some areas' decreases in Twitter activity during a hurricane correlate as strongly with damage as others areas' increases (Samuels et al. 2018a). Researchers do not currently understand what factors could contribute to some populations being represented by social media in an emergency while others simply disappear; however, recent research has suggested that part of the disappearance could be due to sociodemographic vulnerability and the digital divide (Xiao et al. 2015; Zou et al. 2019).

The absence of social media data from vulnerable populations is concerning on two fronts. First, viable and relevant information during a crisis is already in scarce supply, so areas with more information available about the severity of damage are easier and safer targets for resource distribution. In this possible "the squeaky wheel gets the grease" situation, vulnerable populations not contributing to this information stream could receive fewer resources as a result of their lack of data presence. Second, there has been a recent push for emergency information dissemination through social media platforms (Panagiotopoulos et al. 2016; Takahashi et al. 2015). If vulnerable factions are not contributing to social media because they are not using it, then they could also be missing the chance to obtain critical crisis information. It is clear that the decreased presence of vulnerable populations on social media needs to be accounted for in crisis informatics; however, it is not clear how.

The digital divide has been defined as the inequality between "those who have and do not have access to computers and the Internet" (van Dijk 2006). People in lower socioeconomic groups and those associated with vulnerability factors, such as the elderly, are noted to be less likely to have higher levels of internet access and usage (Rogers 2001). If this divide is ascertainable, quantifiable, and able to be delineated prior to a crisis situation, the lack of information coming from these populations can be factored into the usage of social media. Ultimately, if it can be measured, it can be mitigated. However, previous research has only focused on the presence of the digital divide at large scales during and after a crisis situation (Shelton et al. 2014; Xiao et al. 2015; Zou et al. 2019). The relationship between the severity of the divide prior to the identification of a crisis (i.e., prior to the planning period), during the crisis, and beyond the crisis has not been explored. It is possible that a crisis situation introduces new factors that cannot be accounted for in a pre-crisis assessment of social media usage and yet significantly deepen the divide. For instance, lower socioeconomic groups are more likely to lose internet access due to power loss, as internet usage by people in a lower income brackets is often facilitated by free Wi-Fi hotspots located at places of employment or cafes (Khan et al. 2016). During a disaster, those hotspots are no longer available due to closures or travel impedances.

Because of this, the built environment and resilient infrastructure could be a critical piece of how the digital divide could widen during a disaster. The digital divide, as currently defined in disaster research, focuses on those who do not have access to technology in their day-to-day lives. Disaster-induced infrastructural failures could impose technology-access restrictions on vulnerable populations unaffected during normal conditions. Previous work has shown that infrastructure service disruptions caused by a disaster have a diffuse and strong societal impact, and work is being done to delineate risk hotspots prior to disasters (Esmalian et al. 2019). Energy infrastructure, critical to communications networks, are significantly impacted by natural disasters in ways that are still being quantified (Ilbeigi and Dilkina 2018). The uncertainty involved in how crises affect the connection between vulnerable populations and the usage of communication technology is a potential sociocultural "hidden risk" for crisis informatics (Dae Kim 2017).

Understanding how the digital divide is affected by a severe crisis situation, and thus how social factors influence individuals' interactions with technologies in a disaster, is critical to understanding endangered populations' social media representation (Blank 2017). As cities become smarter and more people become more connected to technology, defining the technological data signal from vulnerable populations, especially in disasters, is necessary for understanding what populations could be left behind in our future cities. This research's objective is to identify how a major crisis affects the relationship between vulnerability factors and the prevalence of Twitter activity.

To achieve this objective, the following hypothesis is tested:

H1: Social vulnerability factors decrease the representational capacity of social media more during a crisis period than during a normal period.

This hypothesis is tested in the specific context of Houston, Texas during and after Hurricane Harvey. It is tested through the comparison of principal component regression models constructed using social vulnerability factors and Twitter posts per person (referred to as Twitter activity) in discrete geographic areas in crisis and non-crisis states. This work helps to identify which kinds of people crisis informatics analyses tend to exclude, and whether that exclusion is predictable prior to a crisis. In describing the effect of a crisis on the ability or proclivity of vulnerable populations to Tweet, this research will inform how a crisis influences the severity, reach, and magnitude of any pre-existing digital divide.

4.3 Methods

4.3.1 Social Media Data Acquisition

To address **H1**, the influence of social vulnerability factors on both the Twitter activity during a "normal period" and the activity during a "crisis period" needed to be compared. To achieve this, the researchers streamed and filtered Tweets produced from five weeks prior to the recognition of Hurricane Harvey in the Atlantic (July 9th-August 17th, 2017) and the two-week period following Hurricane Harvey's first landfall in Houston, Texas (August 25th to September 8th, 2017). The first series of dates comprises the "steady state": a normal period during which Twitter users in Houston were not influenced by the oncoming hurricane. The second series comprises the "perturbed state": a period during which Twitter users *were* influenced by the hurricane. The week-long gap between the two (August 18th – August 24th, 2017) accounts for a "transitionary state" during which Twitter users were aware of an incoming crisis and likely modulated their interactions with Twitter despite not being presently affected by physical threats. The use of a five-week steady state was predicated on the recommendation of a four- to six-week steady state by Toepke (Toepke 2018a).

The geolocated Twitter data for the greater metropolitan area of Houston for six weeks prior to and two weeks following Hurricane Harvey's landfall were streamed through the Twitter API (Wang and Taylor, 2015). The raw Tweet count total in the greater Houston metropolitan area for the six weeks prior to the hurricane's landfall was approximately 436,000 Tweets; the raw Tweet count total for the two weeks following landfall was 154,000. Out of concern for the presence of Twitter bots, which are accounts that automatically post Tweets when specific conditions are met, these Tweets were filtered using a mix of keyword analysis and the OSoMe tool Botometer (Indiana University 2018; Toepke 2018b; Yang et al. 2019). Although the comparison of steady and perturbed state precludes the incorporation of textual analysis comparisons, the percentage of Tweets containing one of a series of Hurricane Harvey-related words (such as 'Harvey', 'hurricane', 'flood', 'PrayForHouston', etc.) was analyzed. The percentage of Harvey-related Tweets prior to Hurricane Harvey's first landfall was 6%. This percentage rose to a peak of 62% on August 27th, the date of Hurricane Harvey's second landfall.

Tweets for each day of both the steady and perturbed state were aggregated by day and then plotted in ArcGIS. Heatmaps depicting the geographic distribution of filtered Tweets for one day of the steady state, transitionary state, and perturbed state are presented as **Figure 16a, b,** and **c** respectively.

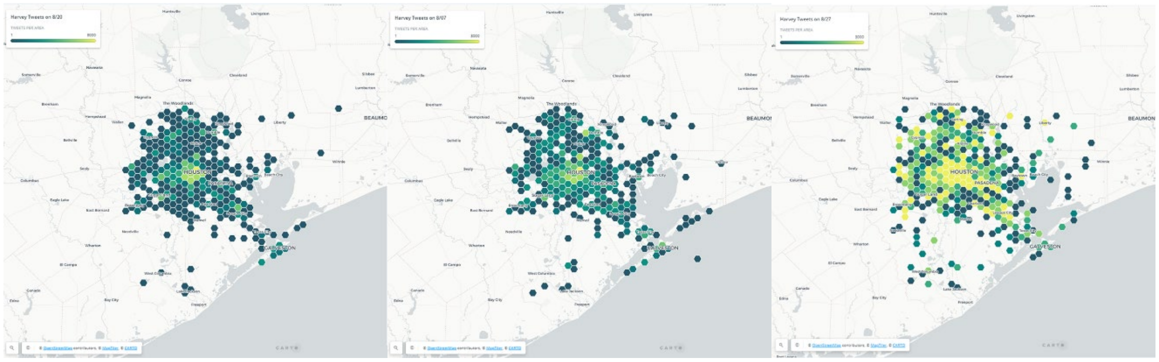

Figure 16a-c. A comparison of the geographical distribution of processed Twitter posts across the city of Houston for different days. Darker green hexagons indicate low Twitter activity, whereas brighter yellow hexagons indicate higher Twitter activity. (Left) is a heatmap for a day of the steady state (August 7th, 2017); (middle) is a heatmap for a day of the transitionary state (August 20th, 2017); and (right) is a heatmap for a day of the perturbed state (August 27th, 2017).

4.3.2 Distributing Demographic Data

Vulnerability indices are based on foundational social research addressing the socioeconomic, mobility, disability, and resource-availability factors that cause discrepancies in the abilities of people to rebuild after disaster. Most of these vulnerability indices are defined at the census tract level and use census data and other social indicator data, such as school performance and community connectedness, to quantify vulnerability. Unfortunately, there are two problems inherent in using census tracts to analyze vulnerability data: the bounds of census tracts are socially biased, and census tracts have such large variations in size that the modifiable areal unit problem (MAUP) can have a significant impact on the results' reliability (Cromley and McLafferty 2002; Nelson and Brewer 2017; Saib et al. 2014). In terms of size, the census tracts within Houston range in size from 0.16 km^2 to 677.20 km^2. This wide range in size can also contribute to data bias in aggregate (Grubesic and Matisziw 2006). Aggregating data in equal-area, uniform shapes can reduce the MAUP's impact on the vulnerability analysis (Nelson and Brewer 2017); however, this required redistributing the census data to a more granular scale.

The geographic information science (GIS) field has historically utilized National Land Cover Database (NLCD) data to increase the granularity of census data with substantial accuracy (Reibel and Agrawal 2007). The NLCD contains a raster file with 30 meters (m) by 30 m cells that have been classified, through satellite imagery, as one of 16 classes. The classification includes four classes of developed land: open space, low intensity, medium intensity, and high intensity. These classifications are determined using the amount of water-permeable and water-impermeable land; thus, the "open space" designation does not necessarily describe areas with no people, but rather areas with a relatively smaller (<20%) amount of concrete-covered land, like suburbs. The raster cells from the 2011 dataset that were classified as "developed" were extracted and clipped to the greater metropolitan Houston Area. Using ArcGIS' Raster to Point function, each of the raster cells were transformed into points located at the center of each cell and spatially joined by count into the census tracts for Houston. Using the counts of each type of NLCD class and the population record for each census tract, a multiple linear regression analysis was performed to determine the expected contribution of each type of land type to the tract's population. The model is presented as Eq. 2.

$$Pop_t = \beta_0 + \beta_1 OpenSpace_t + \beta_2 LowInt_t + \beta_3 MedInt_t + \beta_4 HighInt_t + \epsilon \tag{4}$$

In Eq. 2, Pop_t is the population for a given census tract t; $OpenArea_t$ is the number of NLCD cells described as "Open Space (Developed)" within the census tract; $LowInt_t$ is the number of cells described as "Low Intensity (Developed)" within the tract; $MedInt_t$ is the number of cells described as "Medium Intensity (Developed)"; and $HighInt_t$ is the number described as "High Intensity (Developed)". The results of the regression analysis are presented in **Table 6.** The model has an adjusted R-squared of 0.8317 and a model p-value of <0.001.

Table 6. Multiple linear regression results for the NLCD land class types and the census data.

NLCD Class	Model Coefficient	STD Error	p-value	Re-scaled Coefficient
Open area	0.30	0.03	<0.001***	0.23
Low intensity	0.11	0.07	0.123	0.19
Medium intensity	4.50	0.10	<0.001***	0.58
High intensity	-2.34	0.12	<0.001***	0.01

The model coefficients were used to determine the weighted averages of each land type within each census tract. As shown, the areas of very intense development have a substantial and significant negative contribution to the residential population of the Houston census tracts. Bian and Wilmot encountered similar results in New Orleans, Louisiana in a study using the same technique to study disadvantaged populations impacted by Hurricane Katrina (Bian and Wilmot 2017). Following ground proofing techniques, they assigned a positive but very small assigned coefficient to the highly developed areas of the city. Following their example for the purposes of the population disaggregation, we determined the ratio of each model coefficient to the coefficient range (-2.34 to 4.50) and used this scaled ratio as the re-scaled coefficient. This process ensured a match between the census data population and the population assigned within each tract such that the disaggregation would be at least as accurate. It also preserved the ratio of the magnitude of impact between categories.

The assigned coefficients were employed as weighted averages in Eq. 3 to determine a population quantity to assign to each of the NLCD point shapefiles.

$$Pop_{cell_i} = \frac{\frac{WA_i}{\Sigma WA_{i \to l}} * Pop_{total_{tract}}}{Tract_{count_i}} \tag{5}$$

In Eq. 3, Pop_{cell_i} is the population represented by a single 30mx30m NLCD raster cell point of NLCD type i within a specific census tract; the NLCD types of “Open area”, “Low intensity”, “Medium intensity”, and “High intensity” are represented as $i \to l$; WA_i is the weighted average for the specific land type; $WA_{i \to l}$ is the sum of the weighted averages for each land type; $Pop_{total_{tract}}$ is the total population within the census tract in which the point is located; and $Tract_{count_i}$ is the total number of points of type i within the specific census tract. The resulting dataset was a grid of point files spaced 30 meters apart that could be aggregated into equal-area, uniform areas with less structural bias than census tracts. Although the accuracy of this method cannot be accounted for at a 30mx30m scale, the regression results and the rescaling of the coefficients lend the assurance of reasonable accuracy for larger-scale aggregation.

For the vulnerable population assignment, the Social Vulnerability Index (SVI) developed at the Centers for Disease Control (CDC) was used (Flanagan et al. 2011). This data is available at the census tract level and the margin of error is included within each estimate, allowing for the removal of factors with an error greater than 90% (the Census Bureau standard). These factors made the CDC's SVI ideal for this analysis. To distribute the vulnerability factors, a simple weighted average was used. The percentages of the population ascribing to the SVI factors were multiplied by the identified population of the points produced prior. For example, if a single point had been assigned a population of 0.96 people through Eqs. 1 and 2 and is located in a census tract with a population that is 12% unemployed, that 30mx30m point is noted to represent the equivalent of 0.11 unemployed peoples. Although there are no fractions of people living in Houston, this work is based on the evidence that vulnerable groups are spatially clustered in cities (Cutter and Finch 2008) and, at aggregated scales, is accurate within reasonable error (Bian and Wilmot 2017).

This method was used to develop an estimate for the number of people ascribing to the following 13 vulnerability categories within each area: people without vehicle access; people with limited English skills; minorities; people living in single parent-households; disabled people; people over 65; people under 17; uninsured people; people without a high school diploma; unemployed people; people below the poverty line; people living in crowded homes (defined as those homes with more than 1.5 people in residence per room); and people living in homes within multi-unit complexes.

It should be noted here that the CDC's SVI includes additional factors; however, "persons living within mobile homes" and the "income" factor were not included within this analysis. This decision was made because each of these violated one or more of the assumptions of the analytical techniques used and is discussed further in the **Limitations** section.

4.3.3 *Scalar Aggregation Nets*

The Twitter and demographic data were aggregated into a spatial net composed of uniform hexagons approximately 5 square kilometers (km^2) in area. Using equal-area, uniform polygons can mitigate some of the bias introduced by ZCTAs through the modifiable areal unit problem (Jelinski and Wu 1996). Hexagons are better suited for tiling large geospatial areas because of their scalability and the reduction in sampling bias from edge effects (Carr et al. 1992). There is a constant tradeoff between analytical reliability and data usability when utilizing different scales for social research. For this study, the size of 5 km^2 was chosen based on (1) the area at which the population data was collected (census tracts in Houston have a median size of 4.9 km^2), and (2) prior research identifying minimal tradeoffs in social media analytical reliability at the 5-15 km^2 (Samuels and Taylor 2019b).

Maps of the redistributed census data and a uniform, equal-area hexagonal net are presented as **Figure 17a-b.** One of the critical factors for choosing this size was that, at 5 km^2 the data meets the regression assumption that the variables are normally distributed; the distributions of both the averaged steady state data and the averaged perturbed state data are presented as **Figure 18**.

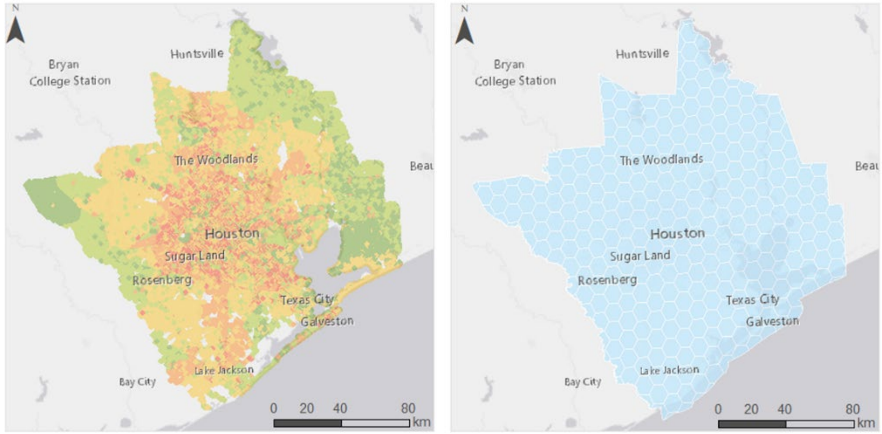

Figure 17a-b. A comparison of the redistributed census data and a uniform, equal-area hexagonal net. (Left) the green areas indicate areas with lower population density, and those in red indicate areas with high population density. (Right) Each hexagon contains 80 km^2; larger hexagons were used in place of those used in the analysis (5 km^2) for improve readability.

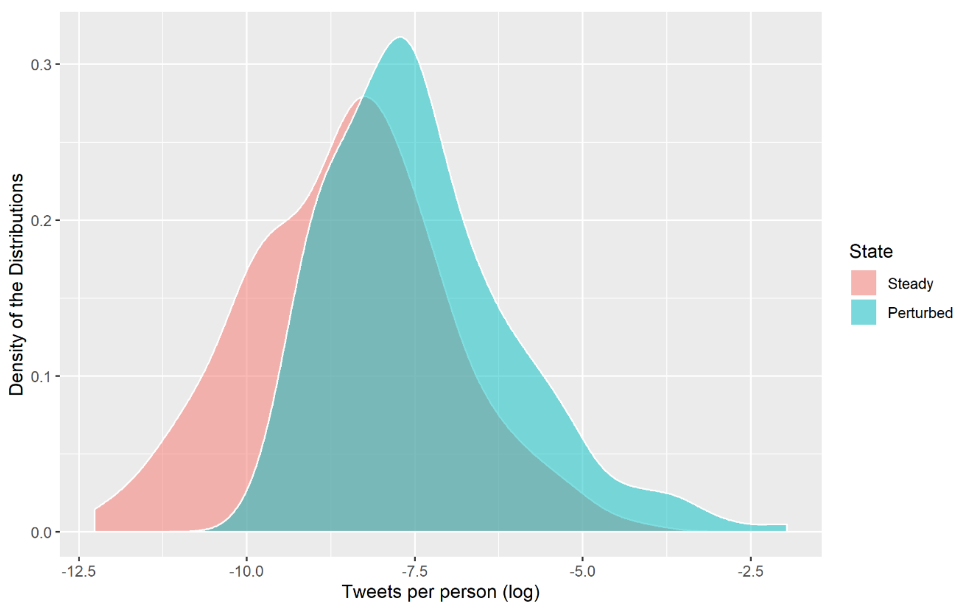

Figure 18. A density graph depicting the distribution of the log of the average Twitter activity across the steady state (pink) and the average Twitter activity across the perturbed state (blue).

The daily Tweets and the redistributed population data were spatially joined (via count and sum respectively) to the 5 km^2 hexagonal net. The Twitter activity for each day was generated by dividing the number of filtered Twitter posts on each day by the estimated number of people residing in each area. The Twitter activity on a "normal" day was generated by averaging the observed Twitter activity values across the defined steady state.

To focus on people in potential danger from the hurricane, Federal Emergency Management Agency (FEMA) Building Level damage assessments were used to identify areas experiencing infrastructural damage. This dataset contains a list of building locations (latitudes and longitudes) and an integer damage rating from 1 to 4 (affected, minor damage, major damage, and destroyed). Other studies have chosen to utilize data from insurance claims to verify or validate relative quantities of hurricane damage or physical storm threat (Kryvasheyeu et al. 2016; Rufat et al. 2019). However, as discussed by Cutter and Finch, "this approach assumes that the most socially vulnerable populations have the most to lose (economically), which is not the case. In correlating property losses with social vulnerability, we would expect an inverse relationship (high social vulnerability; low dollar losses)…" (Cutter and Finch 2008). To mitigate the potential impact of this inverse relationship on the regression model, areas experiencing more or less hurricane damage were identified through the number of FEMA Building Level damage assessments performed in each 5-km^2 area and not insurance data.

Finally, from the three datasets amassed through the methods above (Twitter posts per person, estimated percentages of people described by social vulnerability factors, and the number of FEMA Building Level damage assessments that identified infrastructure damage), areas of interest were extracted from the original 4,849 hexagons. Of those, 1,113 had residential populations as recorded by the census and the disaggregation method. Second, crisis informatics and the relevance of the analysis is dependent on populations that contribute to social media before a storm. Therefore, polygons with zero Twitter activity during the steady state period was removed from the analysis, leaving 452 hexagons. As emergency managers are specifically interested in people in areas strongly impacted by crises, only areas that had an instance of a FEMA Building Level damage assessment of at least "minor damage" were used. The remaining 213 areas are depicted in **Figure 19a-b**. **Figure 19a** depicts the short-listed hexagons color-coded according to the distribution of disabled peoples and overlaid by the FEMA Building Level damage assessments, and **Figure 19b** depicts the distribution of unemployed peoples in the short-listed hexagons.

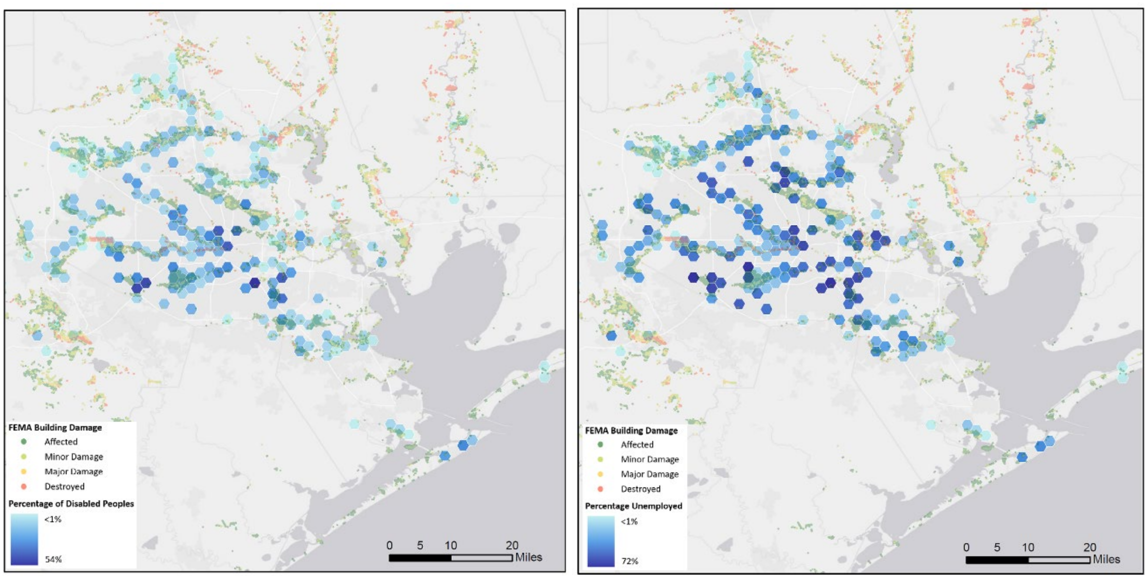

Figure 19a-b. Maps depicting the short-listed hexagons used in the PCR analysis. The hexagons are colored to indicate the relative number of persons within those areas that are described by a vulnerability factor, with a lighter blue indicating fewer vulnerable people and a darker blue indicating more vulnerable people. (Left) shows the distribution of disabled peoples, and (Right) shows the distribution of unemployed peoples. Overlaid across the hexagons is the distribution of FEMA Building Level damage assessments, which are colored on a gradient from dark green to red, with green indicating less damage and red indicating more damage.

4.3.4 Principal Component Regression Analysis

With the population data and Twitter data reaggregated into equal area, uniform polygons, the impact of social vulnerability of Twitter data needed to be assessed. Regression is often utilized to determine the relationship between variables; however, social vulnerability factors tend to be highly collinear. The Pearson correlations of the percentages of people described by the social vulnerability factors of interest are depicted in **Figure 20**. One common method for removing the complication of multicollinearity in vulnerability factors is principal component analysis (PCA).

PCA has been used consistently in vulnerability science to determine the relationships between vulnerability factors and disaster impacts (Cutter and Finch 2008; Khajehei et al. 2020; Lou et al. 2012; de Loyola Hummell et al. 2016; Mavhura et al. 2017). It has also been used in the context of social influences on critical infrastructure issues in identifying the most salient and relevant impactors (Yap et al. 2019). The process transforms a collection of input variables into a series of statistically independent components. Each component is composed of "loadings" that are essentially the contribution of each variable to that component. The loadings are determined through covariance analyses between the variables. The first component that is produced is a vector that describes the most variance possible across the variables; the second component is a vector that is orthogonal to the first and describes the most variance possible of the remaining variability, and so on (Glen et al. 1989). PCA is preferred because it reduces the data's dimensionality (number of considered variables) while preserving the underlying factors that connect the input variables through variation analysis (Abdi and Williams 2010).

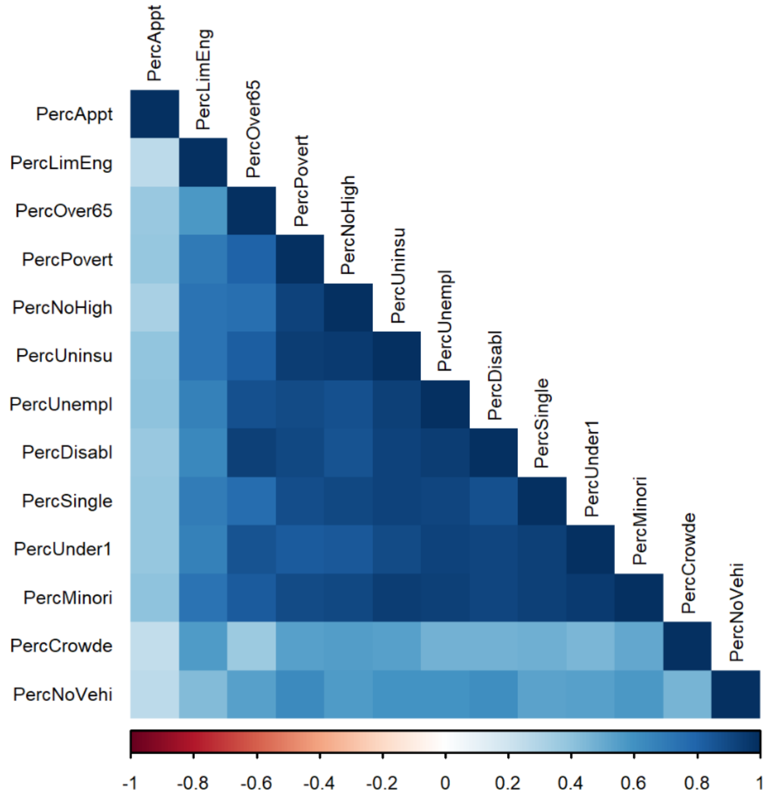

Figure 20. A Pearson correlation matrix displaying the covariance of the sociodemographic factors selected for the vulnerability analysis. The factors included were the percentage of: crowded homes, people in poverty, people with limited English-speaking abilities, no high school diploma, the uninsured, unemployed people, people with physical disabilities, single-parent households, people under 17, people in minority groups, people living in apartments, people over the age of 65, and people without a vehicle.

Following the acquisition of independent principal components that contain the vulnerable populations data, the components most suitable for regression needed to be determined. The number of components that should be used in the regression analysis was determined by the "one-sigma method", which involves calculating the validation plots with the root mean square error of prediction (RMSEP) and identifying the model with the fewest components that is less than one standard error away from the best model (Bro et al. 2008). Per standard principal component regression (PCR) practice (Wehrens et al.

2019), the chosen principal components and the dependent variable (Twitter data) were standardized and scaled. These components were then utilized in a regression analysis described by Eq. 4.

$$TwAct_{state} = \beta_0 + \beta_1 PC_1 + \cdots + \beta_n PC_n \tag{4}$$

In Eq. 4, $TwAct_{state}$ is the Twitter activity (daily Tweets per person) in the temporal state being analyzed. PC_1 is the first principal component, and n is the number of components determined through the one-sigma method described above. The amount of variability in $TwAct_{state}$ for both the steady state average daily Twitter activity and the daily Twitter activity observed in the perturbed state were compared using k-fold cross-validation (CV) with ten segments. This was one of the benefits of using PCR: the regressors were the same in both models. Otherwise, results generated using different response variables would not be explicitly comparable (Glen et al. 1989; Mevik and Wehrens 2015).

Finally, as one of the research objectives for this work was to identify which vulnerability factors contribute more or less to variation in Twitter activity, jackknife validation (similar to bootstrapping) was performed to compare the strength and significance of the input vulnerability variables for each model. Jackknife validation through the R package 'pls' (2.7-2) was chosen to predict the variance of the estimators (Lachenbruch and Mickey 1968).

4.4 Results

The number of components *n* for each model were analyzed first. As this aspect of the PCR process can be quite subjective, graphs depicting the RMSEP of PCR models produced with different quantities of incorporated principal components are depicted in **Figure 21a-c**. Smaller RMSEPs indicate models with less error. **Figure 21a** depicts the RMSEP validation plot for the steady state model; **Figure 21b** depicts the RMSEP validation plot for the date of Hurricane Harvey's first landfall in Houston (August 25th, 2017), and **Figure 21c** shows the plot for Harvey's second, more damaging landfall (August 27th, 2017). These two perturbed state days were chosen as examples due to their heightened impact and thus their greater relevance to emergency management. Based on the number of components that minimized the RMSEP for these models, three components ($n=3$) were used in the PCR model described in Eq. 4 for the model for the first landfall; one component ($n=1$) was used for the PCR model for the second landfall, and three components ($n=3$) were used for the steady state.

Because two of the three depicted models utilized three principal components, the contribution of each of the vulnerability factors to the first three components of the model for August 25th, 2017, for August 27th, 2017, and for the average steady state day are depicted in **Figure 22a-c** respectively. The first principal component (PC1), which describes the greatest amount of variance (80%) and has the greatest influence on the model, is of primary interest in this analysis. Each of the first components for the models are impacted by each vulnerability factor in roughly the same proportion, as would be expected considering their high degree of multicollinearity depicted in **Figure 20.**

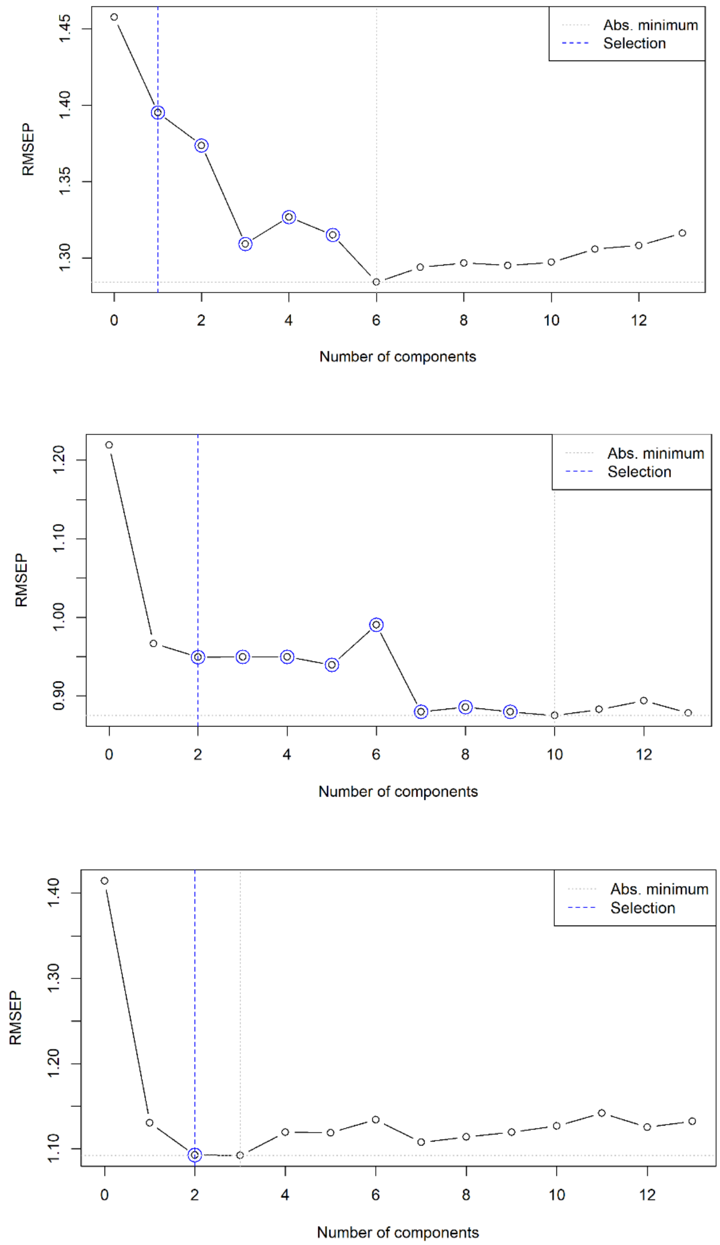

Figure 21a-c. Graphs depicting the relationship between the RMSEP and the number of components included in the PCR model for (top) Twitter activity prediction in the steady state period, (middle) on Hurricane Harvey's first landfall (August 25th, 2017), and (bottom) on the second landfall (August 27th, 2017). The dashed blue vertical lines indicate the number of components selected to be incorporated into the final models through the one-sigma method. The blue circles indicate the one-sigma bands around the RMSEP values. The light gray dotted line shows the absolute minimum.

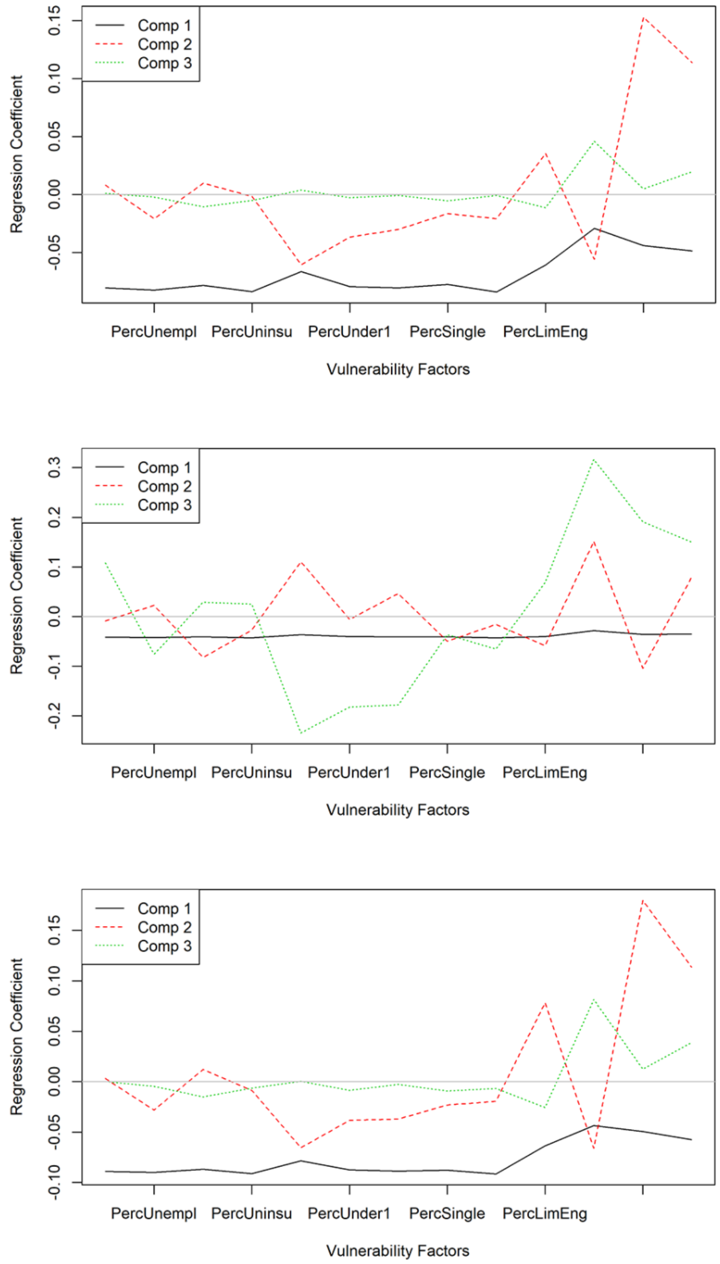

Figure 22a-c. Graphs depicting the contribution of individual vulnerability factors to the first three principal components developed through PCA for (top) Twitter activity prediction in the steady state period, (middle) on Hurricane Harvey's first landfall (August 25th, 2017), and (bottom) on the second landfall (August 27th, 2017). The black lines indicate the weights associated with the separate PC1s; the red dashed lines indicate those for the separate PC2s; and the green dotted lines indicate those for the separate PC3s.

The percentage of variation in the Twitter activity described by the PCR model and the jackknife validation results for the input coefficients for the steady state and both landfall dates are shown in **Table 7.** The amount of variance in Twitter activity described by PC1 is drastically different between steady and perturbed states. Vulnerability factors appear to predict approximately 9% of the variation in Twitter activity during a "typical day"; however, they predict 41% of the variation in Twitter activity when Houston was suffering the worst of Hurricane Harvey's wrath. Additionally, the variables that significantly predict the 9% of Twitter activity during the steady state are limited to the unemployed, the disabled, minorities, and people living in apartments (which contributes positively). This is in direct opposition to the variable coefficients observed in a crisis period, which are almost all extremely significant ($p<0.001$) and negative. This trend is observed more strongly on Hurricane Harvey's second, more destructive landfall (August 27^{th}) than its first (August 25^{th}). The coefficients for each of the vulnerable populations except for minorities and disabled populations observed for the Twitter data on the 27^{th} are stronger and more significantly correlated with a decrease in Twitter activity.

The amount of variance in Twitter activity peaks on the 27^{th} and generally declines afterwards. This trend matches that of the rainfall data for the week following Hurricane Harvey and is shown in **Figure 21**.

Table 7. Variable coefficients derived from the PCR model for the steady state Twitter activity and the two landfall dates for Hurricane Harvey.

Social vulnerability factor[a]	Steady state model coefficient	Perturbed state	
		Aug 25, 2017 model coefficient	Aug 27, 2017 model coefficient
% In Poverty	0.06	-0.08***	-0.09***
% Unemployed	-0.10**	-0.08***	-0.12***
% No High School	-0.09	-0.08***	-0.07**
% Uninsured	-0.05	-0.08***	-0.10***
% Over 65	-0.16	-0.07***	-0.14***
% Under 17	-0.23	-0.08***	-0.13***
% Disabled	-0.17*	-0.08***	-0.13***
% Single Parent Household	-0.13	-0.08***	-0.11***
% Minority	-0.12*	-0.08***	-0.11***
% Limited English	-0.03	-0.06**	0.01
% Apartment	0.44***	-0.03*	-0.11*
% Crowded Homes	0.05	-0.04*	0.13*
% No Vehicle	0.19	-0.05	0.06
% Twitter Activity Variance Explained	**9.7%**	**33%**	**41%**

Note: p-value <0.0001***, < 0.001**, <0.01*, <0.05’

[a]The social vulnerability factors shown above correspond, in order, with the following: people below the poverty line; unemployed people; people without a high school diploma; uninsured people; people over 65; people under 17; disabled people; people living in single parent-households; racial minorities; people with limited English skills; people living in crowded homes; people without vehicle access; and people living in homes within multi-unit complexes.

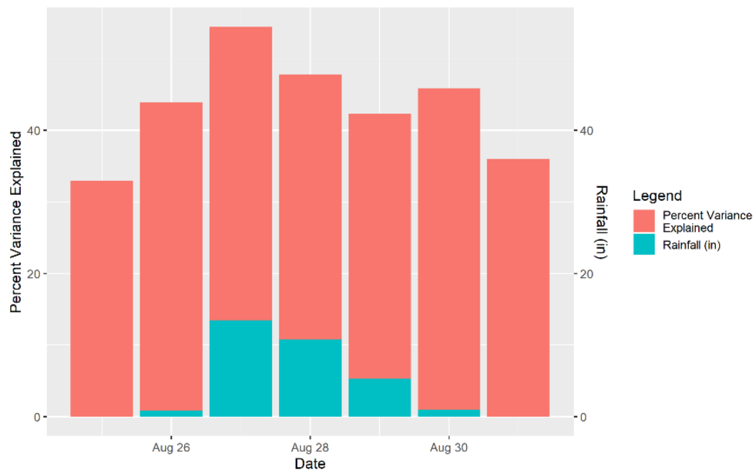

Figure 23. A barplot depicting, in pink, the percentage of variance in Twitter activity explained by the percentage of vulnerable populations within each area, and in blue, the amount of rainfall recorded at the Houston International Airport (NOAA).

4.5 Discussion

These results confirm a distinction in the covariance of populations described by social vulnerability factors and Twitter activity between steady states and perturbed states. They confirm **H1** by showing that the covariance of social vulnerability and per capita Twitter activity is near-negligible during non-crisis periods yet are profound during a crisis. That vulnerable populations tend to contribute less to the social media stream during a hurricane had been identified previously (Wang et al. 2019; Xiao et al. 2015; Zou et al. 2019); however, the influence of the crisis itself on this disparity had not been confirmed with respect to a time period completely external to the crisis (outside of preparation, response, and recovery periods), nor had the severity of that influence been quantified. Without this comparison, the link between a crisis and a worsening digital divide could not be uncovered. The major contribution of this paper is the determination that vulnerable populations decrease their Twitter activity during a major crisis due to factors *separate* from those present during the steady state, such as lack of connectivity.

Thus, although prior research had suggested that crisis informatics can rely less on the representation of vulnerable groups (Zou et al. 2019), this disparity had been attributed to factors quantified by omnipresent, universal metrics like lack of access to phones, computers, or the internet caused by lack of money or stable housing—as opposed to potential infrastructure service failures caused by a hurricane in vulnerable areas. Were this true, crisis informatics could avoid inequity by identifying areas with consistent social media use during a steady state then weighting the influence of social media crisis information higher in those areas. Unfortunately, hurricanes function as a large confounding factor on how vulnerable attributes affect Twitter activity. The predictors of Twitter activity are shown to not be the same between states, and the influence of vulnerability factors on crisis Twitter activity cannot be predicted using steady state Twitter activity. The digital divide likely exists in some form prior to a hurricane; however, the hurricane appears to severely worsen that divide. Accounting for the exact influence of a crisis on access to technology will be more difficult than contextualizing in-crisis Twitter activity by historical Tweeting (as argued for by Chen et al. 2013) or weighting information importance by considering both population and the Odds Ratio (de Albuquerque et al. 2015; Takahashi et al. 2015).

These points are especially poignant when considering how much greater of an impact sociodemographic factors have on Twitter activity when compared to the experience of hurricane damage. Similar to Xiao et al., this research finds that demographic factors are better predictors of Twitter activity in a hurricane than the quantity of hurricane-induced damage (Xiao et al. 2015). What is shown in contradiction to their conclusions, however, is that this disparity is not necessarily a direct result of the digital divide (i.e., pre-crisis lack of access to technology for vulnerable populations). Pre-crisis, variability in the percentage of vulnerable populations can explain less than 10% of the variance in Twitter activity. This indicates that crisis informatics should not assign the blame for the disparity in activity to omnipresent social factors, which might be identifiable, measurable, and thus predictable before a crisis hits. To ascertain what populations are equitably represented by social media produced during a crisis, further research will clearly need to identify and quantify what aspects of a crisis are, in fact, the ones deepening the digital divide.

The noted strong covariance between social vulnerability factors and Twitter activity is also very persistent across time. Research has repeatedly shown that vulnerable populations find it more difficult to recover from disasters, and their recovery takes longer (Comfort et al. 1999; Flanagan et al. 2011; Villegas et al. 2018). This longer recovery time can be attributed to lack of savings, lack of other intangible assets such as community connectivity, weaker building infrastructure, and lack of knowledge or access to post-emergency aid (Fothergill and Peek 2004). Models describing strategies to improve post-disaster insurance strategies emphasize the influence of income level on the tendency of families to insure themselves; the ability of vulnerable families to restore their communication infrastructure could lag far behind others (Eid et al. 2015).

The presented results show a significant but slow decrease in the negative covariance between vulnerability factors and Twitter activity post-landfall and across the recovery period. This shallow slope could indicate that the lack of access to resources noted in vulnerability studies includes hurricane-induced lack of access to internet or other communication technology. Further necessary work includes the comparison of social media access and use recovery rates between vulnerable and less vulnerable populations. This is particularly necessary in the context of online crisis communication. If the hurricane-induced disparity between populations is due to lack of access to technologies, crisis manager communications regarding access to federal aid, shelters, or resource availability could be missed for a substantial period following landfall.

The exact impact of hurricane severity on the divide is also called into question. The presented models show slightly stronger, more significant coefficients for the Twitter activity observed on Harvey's second landfall as compared to its first. Harvey's second landfall generated a profound amount of flooding across the city, and is largely considered to be the period during which the most infrastructural damage and life-threatening situations were caused (Amadeo 2017; Zurich and Global Disaster Preparedness Center 2018). Other crisis informatics studies that have identified geographic data shadows—areas in which social media data was severely diminished or missing during a storm—were performed on Hurricane Sandy (Shelton et al. 2014; Xiao et al. 2015), another exceptionally damaging hurricane even in the recent history of increasingly damaging hurricanes (Gall et al. 2011; Hauer et al. 2016). Stronger hurricanes also have a stronger impact on energy infrastructures, potentially leading to decreased communications access (Ilbeigi and Dilkina 2018; Reed et al. 2010). If the negative influence of vulnerability factors is more prevalent during episodes of increased hurricane damage, it is even more important for the crisis informatics community to determine how to balance the use of social media crisis information against the communities it best represents. Otherwise, social media crisis informatics run the risk of failing vulnerable communities when they might need help and resources the most.

To mitigate that risk, more research is necessary to investigate the relationship between vulnerability and crisis-specific drop-offs in social media usage. Future work should concern itself with further delimiting the effects of specific vulnerability indicators on the relative decrease in available information from specific populations. Although the components generated through PCR contain aspects of distinct factors, the geographic clustering of vulnerable populations (Cutter and Finch 2008; Shelton et al. 2015) and thus their spatial collinearity can make differentiating demographic-specific correlations difficult, particularly at spatial scales smaller than those of data collection. Previous work has shown a slightly larger decrease in crisis data available from populations without access to vehicles, the disabled, and the elderly (Samuels and Taylor 2019a; Zou et al. 2019); however, the interactions between human social media behavior, resource availability, and infrastructural resilience still need to be accounted for.

Previous research into the distribution of vulnerability as it relates to infrastructure, hazard threat, and demographic factors notes that the vulnerability of U.S. cities to hazards, and the resulting variation in recovery time, is highly city-dependent (Borden et al. 2007). The presence or absence of this phenomenon, in more and less vulnerable cities, should be explored. The city-specific context of crisis information and vulnerable populations also informs the potential for a wide variety of influences on the quantity of available information. One of the reasons why Houston was ideal for this analysis is because there were no evacuation orders given to the city itself prior to August 25th, 2017. Officials told citizens to stay home instead of evacuating, thus evacuation had a lesser effect in Houston than it might have in, for instance, Miami as it was hit by Hurricane Irma. In another city, areas with less vulnerable populations may exhibit a sharper decrease in Twitter activity because those populations are more easily able to evacuate (Bian and Wilmot 2017). The city context additionally matters in the context of the previously highlighted data bias against rural populations. The density of people and, separately, the density of Twitter users, need to be incorporated into how, when, and where social media is used for crisis response (Hecht and Stephens 2014)

From a civil engineering perspective, this research isolated areas impacted by infrastructural damage; understanding the influence of the influence of city-wide infrastructure damage and service disruption, especially as related to power and telecommunication coverage, should be explored. A stronger link to the distinct between the influences of infrastructure and human social media behavior should be established in order to help understand how to measure and thus ameliorate disasters' impact. Especially as this research utilized original Twitter postings as its metric for social media involvement, the results cannot specifically inform decisions about the potential gaps in emergency information distribution or risk assessment. Understanding what factors are lessening vulnerable populations' social media involvement will be integral for crisis communication through social media as well (Fan et al. 2019; Lachlan et al. 2016; Olteanu et al. 2015).

4.5.1 Limitations and Future Work

First, although this research intended to highlight an aspect of bias within the use of Twitter data in crisis response, this research is not immune to bias itself. The data is limited to Tweets that are geotagged, which introduces a secondary factor of population bias (Malik et al. 2015). Existing methods for verifying a geotag location (through, for example, the location given in a user's profile) have made only modest changes in location and temporal confidence, and there are substantial demographic biases that are difficult to account for (Jiang et al. 2019). Future work on this subject should consider the combination of multiple location-validation metrics (Grace et al. 2017).

One major limitation of this study is the population distribution technique. First, it was not possible to assess the accuracy of the technique at the 5 km^2 aggregation except through ground-truthing methods. Second, while re-scaling the coefficients ensured that the disaggregated data (if re-aggregated into census tracts) would exactly match the census tract data and that no areas would have a "negative" population, it diminished the geographical variance in population data. These two concerns, while mitigated by the choice of 5 km^2 aggregation areas and the good fit of the regression model, should be explored further. Additionally, in terms of the distribution of the vulnerable populations, a more fine-tuned and nuanced distribution would be necessary for more localized analyses. A more accurate model would redistribute the SVI factors based on additional known vulnerability factors, such as historic redlining, areas of historic segregation, and areas of previous hazard damages. However, that data is not widely or consistently accessible to the authors across the city of Houston and, at the 5 km^2 scale, in unlikely to strongly affect the conclusions presented here. Finally, the NLCD distribution is limited to nighttime and not daytime accuracy. That said, based on the Twitter data text content, many businesses were closed and many people stayed in their homes regardless.

Ultimately, based on the above, research on population disaggregation for crisis research is necessary. Particularly because some localized crises can happen at very small scales (Wurman and Kosiba 2018). In their paper, Wurman and Kosiba theorized that the existence of small-scale vortices could cause some major discrepancies between infrastructural damages would occur at the sub-county level. It is possible that major discrepancies in social media responses reacting to storm damage could vary at small scales, which could not be captured in this paper.

An additional vulnerability factor limitation is the absence of mobile home occupants and income data. Unfortunately, although mobile homes are severely impacted by hurricanes, the data violated two assumptions of the chosen data analysis: lack of outliers and sampling adequacy. The mobile home data for the whole dataset is bimodally distributed, with local maximums at both zero and in the 90^{th} percentile. The number of areas with sufficient mobile home data in our study area did not meet the method's requirements for sampling adequacy. This is likely due to the geographic clustering of mobile homes and trailer parks. Finally, the "income" factor provided by the CDC is the average income for the census tract, and not the "percent of people" utilized for each of the other factors. Although this could theoretically be quantified in the same way by using income intervals, income data is only available as an average and not as a distribution, so the numbers of people at or below a certain income threshold cannot be determined. Fortunately, both of these factors are moderately or strongly correlated with other SVI factors, and their exclusion likely does not significantly impact the results.

In terms of spatial filtration, this research was isolated to areas that sustained infrastructural damage during Hurricane Harvey. This was performed to avoid the false equivalence of increased insurance claims with increased personal threat from the hurricane, but it fails to account for the potential confounding factor of vulnerable populations being more likely to reside within vulnerable infrastructure. The incorporation of the presence of hurricane damage as a presence/absence factor instead of an ordinal one was to mitigate this potential influence.

Another limitation of this work is the temporal granularity of a single day. Recent research has identified the ability of Twitter to detect sub-events (events on more localized space and time scales) through burst and topic analysis (Arachie et al. 2019). Reactions to and complications with sub-events, such as the release of the dams on the Addicks and Barker reservoirs on Buffalo Bayou and the evacuation of the Ben Taub hospital during Hurricane Harvey, can be critical pieces of information for emergency managers. Future research should investigate how the bias against vulnerable populations highlighted here also affect their ability to help identify sub-events, or the topic distribution of those populations' Tweets. Topic modeling of Tweets from vulnerable and non-vulnerable areas across the four stages of a disaster could inform how to better incorporate those areas into social media analyses for emergency management.

Lastly, and perhaps most importantly, it should be noted that this research is unable to isolate the causes, social or technical, of the differences in social media interaction within vulnerable populations. Future work should be focused on distinctions between scalar aggregations and the root causes of the Twitter activity disparity.

4.6 Conclusions

In the midst of increasingly dangerous hurricanes and populous coastal cities, addressing the deficit of reliable, accurate crisis information is a necessity. Social media and volunteered geographic information offer one potential source of information midcrisis; however, equitable stewardship of this data requires understanding who this data best represents. Previous research has shown that this data is not equally representative of all populations in the middle of a crisis, motivating work to delineate how this data can best be used. The research presented in this paper shows that blind usage of social media data will prioritize resource distribution to the least vulnerable in a way that is antithetical to urban resilience. Furthermore, this research establishes that the presence or absence of social media during a steady state cannot quantify the impact a crisis will have on social media usage during a crisis. The major contribution of this paper is evidence that there is something fundamentally different in how vulnerable populations are able to—or want to—use social media during a crisis as opposed to a steady state period.

This drastically complicates how crisis informatics should prioritize social media information. Balancing the usage of new and important forms of information with which populations are most aided by that information is going to be one of the principal ethical questions concerning the pursuit of the smarter, more resilient city. As urban analytics and decision-making begin to utilize more big data produced by the interactions between people and technology, delineating *why* a crisis-specific discrepancy in social media data exists will be critical for equitable data stewardship. Ultimately, emergency responders seeking to use social media data (and any form of humans-as-sensors data) need to incorporate the in-crisis discrepancies between data produced by general and vulnerable populations. In so doing, we can begin the process of bridging these crisis-centric gaps in technological infrastructure instead of deepening the divide.

4.7 Data Availability Statement

Some data, models, or code that support the findings of this study are available from the corresponding author upon reasonable request:

1. The aggregated Twitter count data for any of the days or areas included in the analysis
2. The FEMA Building Level damage assessment data for Houston, TX
3. The code utilized to aggregate the data
4. The code utilized to perform the PCR analysis

Some data, models, or code generated or used during the study are proprietary or confidential in nature and may only be provided with restrictions (e.g. anonymized data):

1. The Twitter data utilized herein has personally identifiable information and cannot be provided as it was streamed from the Twitter API. Twitter, Inc. has additionally restricted the amount and kind of data that can be shared between end users. Therefore, individual Tweets cannot be provided; only aggregated (and thus anonymized) data can be provided.

4.8 Acknowledgements

This material is based upon work supported by the National Science Foundation under Division of Information and Intelligent Systems (IIS) Grant No. 1760645. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

CHAPTER 5. CONTRIBUTIONS

Within the body of work presented herein, I have tested the boundaries of social media data for usage in crisis response. I have sought to understand what can be learned through an absence of data; what influence scale has on analytical conclusions; and how disadvantaged populations are underserved in social media analyses. As the potential utility of social media continues to be explored by researchers and emergency managers, it is imperative that the equity and limitations of that data are not only acknowledged but incorporated.

To assist with taking the first steps down that path, I began my doctoral research by examining how sudden social media silence could be an indication of severe instead of minimal harm. Using steady state and perturbed state analyses, I showed that incorporating social media activity deviation instead of only social media bursts could better identify areas of infrastructural damage. In my second study, I showed how changes in the scale at which social media data is analyzed strongly affects how many areas show statistically reliable quantities of data, the level of correlation between Twitter activity and infrastructural damage, and whether bursts of activity or a sudden absence of it is more common during a crisis. In my last study, I showed that social vulnerability factors strongly negatively influence the amount of data produced in areas with a high degree of socially vulnerable populations, and that this influence is much stronger in a crisis state than in a normal state. The results of these studies were intended to be applied towards both social media crisis applications developers and emergency managers seeking to make sense and correctly interpret the social media data generated during crises. In the following sections, I describe the contributions of each study and opportunities for future research based on those contributions.

5.1 Silence of the Tweets

I began my research just as Hurricane Harvey was wreaking havoc on Houston. As I started to explore the innovative research being performed using social media data, I first encountered the research that Shelton et al. had performed on Twitter data from Hurricane Sandy (Shelton et al. 2014). Although Shelton and his team had focused on the geography of Twitter and the spatial relationships in the data, they noted in a supplemental analysis that Staten Island—which contained 50% of the deaths incurred by Hurricane Sandy—had almost a complete lack of Twitter posts. Shelton referred to blank spots such as this as "data shadows" and commented on the lack of ubiquity of social media data. Due to the disparity between this observation and the often-touted positive correlation between Twitter activity and infrastructure damage (Kryvasheyeu et al. 2016), I initially sought to confirm the theory that sudden social media silence like that on Staten Island was not an outlier but rather a reliable indicator of hurricane damage. Within *Silence of the Tweets*, I show that the absolute deviation in Twitter activity from a steady state has a stronger, more positive, statistically significant correlation with damage as compared to the raw deviation. This confirms the theory that absence of evidence of hazardous events on social media is not evidence of absence; indeed, that the opposite is more likely.

Although I cannot advise a crisis application taking the equations and methods I developed straight out of the paper and attempting to apply it directly to existing applications, I believe this is proof enough of concept that organizations using social media data should develop comparative baseline models. By this I mean that areas with substantial amounts of social media in a non-crisis state that go silent should be investigated during a crisis instead of ignored, or considered areas in which the humans-as-sensors network is down, so to speak. Many of the crisis map applications currently have multiple different views and data layers than can be viewed by emergency managers (Chen et al. 2016; Imran et al. 2014); I believe this research is a strong argument for including “deviation from reference state” as one of the layers. Additionally, I believe the data show a trend of increasing amounts and significance of social media silence in cities which experienced greater amounts of hurricane damage. If social media silence becomes more significant with increasing hurricane damage, and if hurricane damage is expected to increase (Gall et al. 2011), then this implies an increasing need to consider the impacts of the silence I identified.

Through this research, I have identified two main promising future research paths. The first is to better understand when and how social media silence can best be an indicator of a crisis event. This research provides proof of concept and theory, but it does not provide a directly translatable method of identifying crises through social media silence. The second is understanding, a-priori, what factors influence the prevalence of silence in some areas versus excited bursts of Twitter activity in others. Which of the drop-offs are due to failures in urban resilience, and which are due solely to intentional human behavioral choices? Interrupted network access could have been caused by power outages, a lack of access through a paid mobile network, or through lack of access to external networks. Part of the socioeconomic facet of that question was explored in Chapter 4, but there are still many aspects of socioeconomic vulnerability as it relates to crisis informatics that have not been explored.

This research paper contributes the confirmation that social media data shadows during a crisis are necessary components to consider in crisis management strategies, and that sudden and severe decreases in social media silence are more likely to indicate danger than not. Ultimately, to use humans-as-sensors data in the most equitable and actionable way, crisis researchers need to listen for both the *sound* and the *silence* of the Tweets.

5.2 Tipping the Scales

Following the identification of social media silence as a likely indicator of a local hazard, I began investigating why absences of data have been understudied and unutilized. The point of departure for *Silence of the Tweets* was the Shelton et al. paper, so I continued to review both his lab's papers and other social media studies in the field of critical GIS. I encountered a decades-old problem in GIS called the Modifiable Areal Unit Problem (MAUP), which is essentially an aspect of Simpson's paradox when studying geographic information (Wagner 1982). Averaging across two groups—or aggregating across large areas—can create a trend that is the opposite of what the trend is when the groups are considered in isolate (Jelinski and Wu 1996; Nelson and Brewer 2017). I theorized that the spots of social media silence that I had observed in my first study had been diluted by differences in scalar aggregation in other studies.

Part of this theory was founded in the previously-recognized disparity in urban and rural social media informatics (Hecht and Stephens 2014; Johnson et al. 2017). Most social media studies operate using social boundaries, which are often delineated by population sizes or political boundaries. When using either census tracts or ZCTAs, there is less aggregation at the population-dense centers of cities and more aggregation towards the larger rural tracts on the peripheries. Additionally, the field of critical GIS strongly recommends place-based algorithms and analyses that incorporate community history (Thatcher et al. 2015). Previous research had identified differences in social media correlation with crisis events at the census tract and county scale (Kryvasheyeu et al. 2016), and further research had theorized that larger areas of analysis would have a stronger connection between social media data and crisis events (Chen et al. 2013). However, the actual relationship between analytical scale and social media analyses had not been defined, and the consequences of using non-uniform scales had not been explained in the context of either information availability or crisis detection.

I modified the spatial association of scalable hexagons (SASH) technique developed originally for the field of landscape ecology to assess the effect of scale on various metrics used in the field of crisis informatics to identify crises. These were: bursts of extreme Twitter activity, sudden Twitter activity silence, the correlation between Twitter activity and infrastructure damage, and the geographic area with statistically sufficient quantities of Twitter data for analysis. I identified several potential power law relationships between these factors that are critical to current crisis-identification methods and rapid damage-assessment metrics. In doing so, I contribute several undiscussed tradeoffs between scale and the value of social media data. Decreasing scale improves location specificity and reduces the influence of the MAUP; however, decreasing scale also reduces the statistical robustness of data contextualization, increases the percentage of areas exhibiting extreme social media behaviors, and, perhaps most importantly, exponentially decreases the correlation between social media activity and damage. The results show that the negative influence of these tradeoffs is minimized between 5-15 km^2.

In terms of research applications utilized by emergency managers, again, I cannot immediately recommend that all research applications employ 5 km^2 hexagons to aggregate Twitter data. Part of this is the relative inaccuracy of disaggregated population at larger scales; the other part is that scale likely has different effects on different forms of crises, just as "silence" did. That said, this technique could be pulled directly from this research to produce the reference state discussed in **5.1**. Additionally, in the context of crisis applications, running either *spatial clustering* or *spatiotemporal aggregation* analyses at multiple different scales is necessary for full data contextualization. Beyond that, I believe further research is necessary to carefully delineate the conclusions that can be drawn from social media analyses at the local scale and at the broader scales of a major disaster. For emergency managers, knowing the location, magnitude, and severity of a crisis is paramount. Within this research, I show that how location and magnitude are defined in our algorithms strongly influences the severity determined through social media. Additionally, this influence is likely further disadvantaging rural and vulnerable populations in existing methodology. With our sights set on more equitable and actionable analyses, we must further incorporate geographic scale into how we interpret social media activity.

5.3 Deepening the Divide

Finally, I began to explore in more depth my theory that socioeconomic vulnerabilities were a driving factor in the presence and importance of social media silence.

If this were the case, usage of social media data without contextualization would prioritize resource distribution to the least vulnerable instead of the most. I additionally wanted to test whether a decrease in social media presence in areas with more vulnerable populations could be predicted from the pre-crisis period. Recent research into vulnerable populations has shown that the poorest populations—and poorest countries—are going to be hit first and hardest by the effects of climate change (Schiermeier 2018). Thus, the way we process information needs to be more than peripherally aware of the differences in digital capabilities. Previous research had noted the existence of the "digital divide" in the context of social media during crises: vulnerable populations contribute comparatively less to social media streams than less vulnerable populations during a disaster (Zou et al. 2018). Considering my position arguing for cities to develop reference social media states to which disaster data could be compared, I wanted to understand whether or not the relatively smaller contribution of vulnerable populations during a disaster could be identified in a steady state. If it could, then the digital divide noted by Zou et al. would not be disaster-specific, and the reference state could utilized as a yardstick by which to measure the impact of the crisis on vulnerable populations. If there were a greater disparity between the data observed in vulnerable populations and non-vulnerable populations in a disaster state than a normal state, however, that would be ample evidence of a crisis-specific phenomenon that directly and specifically affects vulnerable populations' contribution to the social media data stream.

The major contribution of this paper is evidence that there is something fundamentally different in how vulnerable populations are able to—or want to—use social media during a crisis as opposed to a steady state period. I show that vulnerability factors have a significant negative influence on social media activity during a hurricane but not during the steady state. This indicates that vulnerable populations decrease their social media usage during a hurricane in a way that less vulnerable populations do not, and that the presence or absence of social media during a steady state cannot quantify the impact a crisis will have on their social media usage. As the crisis informatics community continues to evolve, and as emergency responders are increasingly monitoring social media during a crisis (Murthy and Gross 2017), this consideration of how well the existing data represents different populations will be critical to equity. Areas with higher vulnerability scores have already been shown to be more poorly served by existing emergency response services like hurricane evacuation (Bian and Wilmot 2017); it would be a vast disservice to these populations by continuing this trend in services distributed due to needs identified through social media analyses.

In terms of future work, previous research into the distribution of vulnerability as it relates to infrastructure, hazard impact, and demographic factors notes that the vulnerability of U.S. cities to hazards, and the resulting variation in recovery time, is highly city-dependent (Borden et al. 2007). The presence or absence of this phenomenon, in more and less vulnerable cities, should be explored. The city-specific context of crisis information and vulnerable populations also informs the potential for a wide variety of influences on the quantity of available information. As it stands, this research is unable to isolate the causes, social or technical, of the differences in social media interaction within in vulnerable populations. Because of that, the influence of city-wide infrastructure damage and service disruption, especially as related to power and telecommunication coverage, should be explored. Understanding what factors are lessening vulnerable populations' social media involvement will be integral for crisis communication through social media as well (Fan et al. 2019; Lachlan et al. 2016; Olteanu et al. 2015).

For crisis researchers and emergency managers, this complicates how crisis informatics should prioritize social media information. Balancing the usage of new and important forms of information with which populations are most aided by that information is going to be one of the principal ethical questions concerning the pursuit of the smarter, more resilient city. As urban analytics and decision-makers begin to utilize more big data produced by the interactions between people and technology, delineating *why* a crisis-specific discrepancy in social media data exists will be critical for equitable data stewardship. This could be done through a-priori recognition of areas with high degrees of social vulnerability or through continued assessment of the factors that might be causing the targeted decreases in social media activity from vulnerable populations. Ultimately, emergency responders seeking to use social media data (and any form of humans-as-sensors data) need to incorporate the in-crisis discrepancies between data produced by general and vulnerable populations. In so doing, we can begin the process of bridging these crisis-centric gaps in technological infrastructure instead of deepening the divide.

CHAPTER 6. CONCLUSION

Social media data is a growing, constantly-evolving source of community-based information that has strong potential for use in emergency management. As urban development increases in coastal regions and as climate change increases disaster-risk in those regions, improving the quantity and quality of the data available for crisis response is becoming more important. However, even as crisis management applications using social media data continue to proliferate, it is equally important for us to consider the boundaries of who social media data is capable of representing during a crisis. Within the research presented in this dissertation, I show that areas where there has been a sharp decrease in Twitter activity are more likely to have larger amounts of infrastructural damage than less; that the geographic boundaries of existing analyses can strongly influence the conclusions of those analyses in ways detrimental to people in less-populated areas; and that crises inhibit vulnerable populations' interactions with social media beyond the limitations faced by those populations during normal periods. These studies incorporate methodologies from critical GIS, emergency management, spatial analytics, and information management to produce both warnings and suggestions for utilizing non-traditional forms of crisis information. As citizens increase their contributions to social media data, and as emergency managers look more and more towards using it, understanding how social media data can be equitably and actionably deployed is a necessary step towards creating smarter, more resilient cities.

REFERENCES

Abdi, H., and Williams, L. J. (2010). “Principal component analysis.” *Wiley Interdisciplinary Reviews: Computational Statistics*, 2(4), 433–459.

Adachi, S. A., Nishizawa, S., Yoshida, R., Yamaura, T., Ando, K., Yashiro, H., Kajikawa, Y., and Tomita, H. (2017). “Contributions of changes in climatology and perturbation and the resulting nonlinearity to regional climate change.” *Nature Communications*, Springer US, 8(1).

de Albuquerque, J. P., Herfort, B., Brenning, A., and Zipf, A. (2015). “A geographic approach for combining social media and authoritative data towards identifying useful information for disaster management.” *International Journal of Geographical Information Science*, 29(4), 667–689.

Allen, K. (2006). “Community-based disaster preparedness and climate adaptation: local capacity building in the Philippines.” *Disasters*, 30(1), 81–101.

Amadeo, K. (2017). “Hurricane Harvey Facts, Damage and Costs What Made Harvey So Devastating.” *The Balance*, 1–7.

Arachie, C., Gaur, M., Anzaroot, S., Groves, W., Zhang, K., and Jaimes, A. (2019). “Unsupervised Detection of Sub-events in Large Scale Disasters.”

Ashktorab, Z., Brown, C., Nandi, M., and Culotta, A. (2014). “Tweedr: Mining Twitter to Inform Disaster Response.” *Proceedings of the 11th International ISCRAM Conference*, 11(May), 354–358.

Ashley, W. S., Strader, S., Rosencrants, T., and Krmenec, A. J. (2014). “Spatiotemporal changes in tornado hazard exposure: The case of the expanding bull’s-eye effect in chicago, illinois.” *Weather, Climate, and Society*, 6(2), 175–193.

Aubrecht, C., Özceylan Aubrecht, D., Ungar, J., Freire, S., and Steinnocher, K. (2017). “VGDI – Advancing the Concept: Volunteered Geo-Dynamic Information and its Benefits for Population Dynamics Modeling.” *Transactions in GIS*, 21(2), 253–276.

Bian, L., and Butler, R. (1999). “Comparing effects of aggregation methods on statistical and spatial properties of simulated spatial data.” *Photogrammetric Engineering and*

Remote Sensing, 65(1), 73–84.

Bian, R., and Wilmot, C. G. (2017). "Measuring the vulnerability of disadvantaged populations during hurricane evacuation." *Natural Hazards*, Springer Netherlands, 85(2), 691–707.

Blank, G. (2017). "The Digital Divide Among Twitter Users and Its Implications for Social Research." *Social Science Computer Review*, 35(6), 679–697.

Blumenstock, J. (2018). "Don't forget people in the use of big data for development." *Nature News*.

Borden, K. A., Schmidtlein, M. C., Emrich, C. T., Piegorsch, W. W., and Cutter, S. L. (2007). "Vulnerability of U.S. Cities to Environmental Hazards." *Journal of Homeland Security and Emergency Management*, 4(2).

Boyd, D., and Crawford, K. (2012). "Critical Questions for Big Data." *Information, Communication & Society*, 15(5), 662–679.

Bro, R., Kjeldahl, K., Smilde, A. K., and Kiers, H. A. L. (2008). "Cross-validation of component models: A critical look at current methods." *Analytical and Bioanalytical Chemistry*.

Carr, D., Olsen, A., and White, D. (1992). "Hexagon Mosaic Maps for Display of Univariate and Bivariate Geographical Data." *Cartography and Geographic Information Systems*, 19(4), 228–236.

Chen, C., Neal, D., and Zhou, M. (2013). "Understanding the evolution of a disaster-a Framework for Assessing Crisis in a System Environment (FACSE)." *Natural Hazards*, 65(1), 407–422.

Chen, X., Elmes, G., Ye, X., and Chang, J. (2016). "Implementing a real-time Twitter-based system for resource dispatch in disaster management." *GeoJournal*, Springer Netherlands, 81(6), 863–873.

Choe, S., Park, J., Han, S., Park, J., and Yun, H. (2017). "A study on the real-time management and monitoring process for recovery resources using Internet of Things." *International Research Journal of Engineering and Technology (IRJET)*, 4(3), 2634–2639.

Clauset, A., Shalizi, C. R., and Newman, M. E. J. (2007). "Power-law distributions in empirical data." 51(4), 661–703.

Comfort, L., Wisner, B., Cutter, S., Pulwarty, R., Hewitt, K., Oliver-Smith, A., Wiener, J., Fordham, M., Peacock, W., and Krimgold, F. (1999). "Reframing disaster policy: The global evolution of vulnerable communities." *Environmental Hazards*, 1(1), 39–44.

Cromley, E., and McLafferty, S. (2002). *GIS and public health*. The Guilford Press, New York, New York, USA.

Cutter, S. L., Barnes, L., Berry, M., Burton, C., Evans, E., Tate, E., and Webb, J. (2008). "A place-based model for understanding community resilience to natural disasters." *Global Environmental Change*, 18(4), 598–606.

Cutter, S. L., and Finch, C. (2008). "Temporal and spatial changes in social vulnerability to natural hazards." *Proceedings of the National Academy of Sciences*, 105(7), 2301–2306.

Dae Kim, S. (2017). "Understanding Hidden Risks from Disasters: Cases of Hurricane Katrina and Fukushima Nuclear Meltdown." *Journal of Management in Engineering*, 33(5), 1–4.

David, C. C., Ong, J. C., and Legara, E. F. T. (2016). "Tweeting supertyphoon Haiyan: Evolving functions of twitter during and after a disaster event." *PLoS ONE*.

van Dijk, J. A. G. M. (2006). "Digital divide research, achievements and shortcomings." *Poetics*.

Eckle, M., and de Albuquerque, J. P. (2015). "Quality Assessment of Remote Mapping in OpenStreetMap for Disaster Management Purposes." *Proceedings of the ISCRAM 2015 Conference - Kristiansand, May 24-27*, (August 2016), 1–8.

Eid, M. S., El-Adaway, I. H., and Coatney, K. T. (2015). "Evolutionary stable strategy for postdisaster insurance: Game theory approach." *Journal of Management in Engineering*, 31(6), 1–9.

Esmalian, A., Ramaswamy, M., Rasoulkhani, K., and Mostafavi, A. (2019). "Agent-Based Modeling Framework for Simulation of Societal Impacts of Infrastructure Service Disruptions during Disasters." *Computing in Civil Engineering 2019*, American

Society of Civil Engineers, Reston, VA, 16–23.

Fan, C., Jiang, Y., and Mostafavi, A. (2019). “Seeding Strategies in Online Social Networks for Improving Information Dissemination of Built Environment Disruptions in Disasters.” *Computing in Civil Engineering 2019*, American Society of Civil Engineers, Atlanta, GA, 487–494.

Fan, C., Jiang, Y., and Mostafavi, A. (2020). “Social Sensing in Disaster City Digital Twin: Integrated Textual-Visual-Geo Framework for Situational Awareness during Built Environment Disruptions.” *Journal of Management in Engineering*, 36(3), 1–13.

Fan, C., and Mostafavi, A. (2019). “A graph-based method for social sensing of infrastructure disruptions in disasters.” *Computer-Aided Civil and Infrastructure Engineering*, (May), 1–16.

Flanagan, B. E., Gregory, E. W., Hallisey, E. J., Heitgerd, J. L., and Lewis, B. (2011). “A Social Vulnerability Index for Disaster Management.” *Journal of Homeland Security and Emergency Management*, 8(1).

Fothergill, A., and Peek, L. A. (2004). “Poverty and disasters in the United States: A review of recent sociological findings.” *Natural Hazards*, 32(1), 89–110.

Fotheringham, A. S., and Wong, D. W. S. (1991). “The Modifiable Areal Unit Problem in Multivariate Statistical Analysis.” *Environment and Planning A*, 23(7), 1025–1044.

Gall, M., Borden, K. A., Emrich, C. T., and Cutter, S. L. (2011). “The unsustainable trend of natural hazard losses in the United States.” *Sustainability*, 3(11), 2157–2181.

Gandomi, A., and Haider, M. (2015). “Beyond the hype: Big data concepts, methods, and analytics.” *International Journal of Information Management*, Elsevier Ltd, 35(2), 137–144.

Gao, H., Barbier, G., and Goolsby, R. (2011). “IEEE Intelligent Systems Harnessing the Crowdsourcing Power of Social Media for Disaster Relief.” *IEEE Intelligent Systems*.

Glen, W. G., Dunn, W. J., and Scott, D. R. (1989). “Principal components analysis and partial least squares regression.” *Tetrahedron Computer Methodology*, 2(6), 349–376.

Grubesic, T. H., and Matisziw, T. C. (2006). "On the use of ZIP codes and ZIP code tabulation areas (ZCTAs) for the spatial analysis of epidemiological data." *International Journal of Health Geographics*, 5, 1–15.

Guan, X., and Chen, C. (2014). "Using social media data to understand and assess disasters." *Natural Hazards*, 74(2), 837–850.

Hauer, M. E., Evans, J. M., and Mishra, D. R. (2016). "Millions projected to be at risk from sea-level rise in the continental United States." *Nature Climate Change*, 6(7), 691–695.

Hecht, B., and Stephens, M. (2014). "A tale of cities: Urban biases in volunteered geographic information." *Proceedings of the 8th International Conference on Weblogs and Social Media, ICWSM 2014*, 197–205.

Hiltz, S. R., and Kushma, J. (2014). "Use of Social Media by U . S . Public Sector Emergency Managers : Barriers and Wish Lists." *Proceedings of the 11th International ISCRAM Conference*, (May), 602–611.

"Houston, Texas Population 2018." (2018). *World Population Review*, <http://worldpopulationreview.com/us-cities/houston-population/>.

Ilbeigi, M., and Dilkina, B. (2018). "Statistical Approach to Quantifying the Destructive Impact of Natural Disasters on Petroleum Infrastructures." *Journal of Management in Engineering*, 34(1), 1–11.

Imran, M., Castillo, C., Diaz, F., and Vieweg, S. (2015). "Processing Social Media Messages in Mass Emergency." *ACM Computing Surveys*, 47(4), 1–38.

Imran, M., Castillo, C., Lucas, J., Meier, P., and Vieweg, S. (2014). "AIDR: Artificial intelligence for disaster response." *Proceedings of the companion publication of the 23rd international conference on World wide web companion*, (October), 159–162.

Indiana University. (2018). "Botometer® by OSoMe." *Observatory on Social Media (OSoMe)*.

Jelinski, D. E., and Wu, J. (1996). "The modifiable areal unit problem and implications for landscape ecology." *Landscape Ecology*, 11(3), 129–140.

Jennex, M. E. (2012). “Social Media – Viable for Crisis Response?” *International Journal of Information Systems for Crisis Response and Management*, 4(2), 53–67.

Jessop, B., Brenner, N., and Jones, M. S. (2008). “Theorizing sociospatial relations.” *Environment and Planning D: Society and Space*, 26(3), 389–401.

Jiang, B. (2018). *Trends in Spatial Analysis and Modelling*. *Geotechnologies and the Environment*, Geotechnologies and the Environment, (M. Behnisch and G. Meinel, eds.), Springer International Publishing, Cham.

Jiang, Y., Li, Z., and Ye, X. (2019). “Understanding demographic and socioeconomic biases of geotagged Twitter users at the county level.” *Cartography and Geographic Information Science*, Taylor & Francis, 46(3), 228–242.

Johnson, I., McMahon, C., Schöning, J., and Hecht, B. (2017). “The effect of population and ‘Structural’ biases on social media-based algorithms - A case study in geolocation inference across the urban-rural spectrum.” *Conference on Human Factors in Computing Systems - Proceedings*, 2017-May, 1167–1178.

Jongman, B., Wagemaker, J., Romero, B., and de Perez, E. (2015). “Early Flood Detection for Rapid Humanitarian Response: Harnessing Near Real-Time Satellite and Twitter Signals.” *ISPRS International Journal of Geo-Information*, 4(4), 2246–2266.

Kendall, M. G. (1938). “A New Measure of Rank Correlation.” *Biometrika*.

“Kendall Rank Correlation Coefficient.” (2008). *The Concise Encyclopedia of Statistics*.

Khajehei, S., Ahmadalipour, A., Shao, W., and Moradkhani, H. (2020). “A Place-based Assessment of Flash Flood Hazard and Vulnerability in the Contiguous United States.” *Scientific Reports*, 10(1), 1–12.

Khan, A. S., Rahman, A. ur, and Qazi, L. T. (2016). “The Relationship Between Internet Usage, Socioeconomic Status, Subjective Health and Social Status.” *Business & Economic Review*, 8(Special Edition), 67–82.

Kryvasheyeu, Y., Chen, H., Moro, E., Van Hentenryck, P., and Cebrian, M. (2015). “Performance of Social Network Sensors during Hurricane Sandy.” *PLOS ONE*, (L. A. N. Amaral, ed.), 10(2), e0117288.

Kryvasheyeu, Y., Chen, H., Obradovich, N., Moro, E., Van Hentenryck, P., Fowler, J., and Cebrian, M. (2016). “Rapid assessment of disaster damage using social media activity.” *Science Advances*, 2(3), e1500779.

Lachenbruch, P. A., and Mickey, M. R. (1968). “Estimation of Error Rates in Discriminant Analysis.” *Technometrics*.

Lachlan, K. A., Spence, P. R., Lin, X., Najarian, K., and Del Greco, M. (2016). “Social media and crisis management: CERC, search strategies, and Twitter content.” *Computers in Human Behavior*, Elsevier Ltd, 54, 647–652.

Lieberman-Cribbin, W., Liu, B., Schneider, S., Schwartz, R., and Taioli, E. (2017). “Self-reported and FEMA flood exposure assessment after hurricane sandy: Association with mental health outcomes.” *PLoS ONE*, 12(1), 1–15.

Lou, W. ping, Chen, H. yan, Qiu, X. fa, Tang, Q. yi, and Zheng, F. (2012). “Assessment of economic losses from tropical cyclone disasters based on PCA-BP.” *Natural Hazards*, 60(3), 819–829.

Lovejoy, K., Waters, R. D., and Saxton, G. D. (2012). “Engaging stakeholders through Twitter: How nonprofit organizations are getting more out of 140 characters or less.” *Public Relations Review*, Elsevier Inc., 38(2), 313–318.

de Loyola Hummell, B. M., Cutter, S. L., and Emrich, C. T. (2016). “Social Vulnerability to Natural Hazards in Brazil.” *International Journal of Disaster Risk Science*, Beijing Normal University Press, 7(2), 111–122.

Malik, M. M., Lamba, H., Nakos, C., and Pfeffer, J. (2015). “Population bias in geotagged tweets.” *AAAI Workshop - Technical Report*, WS-15-18, 18–27.

Marzuoli, A., and Liu, F. (2019). “A data-driven impact evaluation of Hurricane Harvey from mobile phone data.” *Proceedings - 2018 IEEE International Conference on Big Data, Big Data 2018*, IEEE, 10, 3442–3451.

Mason, A. M., Drew, S., and Weaver, D. (2017). “Managing Crisis-induced uncertainty: First responder experiences from the 2011 Joplin-Duquesne Tornado.” *International Journal of Disaster Risk Reduction*, Elsevier Ltd, 23(April), 231–237.

Massey, F. J. (1951). “The Kolmogorov-Smirnov Test for Goodness of Fit.” *Journal of the*

American Statistical Association.

Mavhura, E., Manyena, B., and Collins, A. E. (2017). "An approach for measuring social vulnerability in context: The case of flood hazards in Muzarabani district, Zimbabwe." *Geoforum*, Elsevier, 86(October), 103–117.

Mevik, B., and Wehrens, R. (2015). "Introduction to the pls Package." *Help section of the "pls" package of RStudio software*, (Section 7), 1–23.

Milliner, C., Materna, K., Bürgmann, R., Fu, Y., Moore, A. W., Bekaert, D., Adhikari, S., and Argus, D. F. (2018). "Tracking the weight of Hurricane Harvey's stormwater using GPS data." *Science Advances*, 4(9).

Mislove, A., Lehmann, S., Ahn, Y., Onnela, J., and Rosenquist, J. N. (2011). "Understanding the Demographics of Twitter Users." *Artificial Intelligence*, 554–557.

Morstatter, F., and Liu, H. (2017). "Discovering, assessing, and mitigating data bias in social media." *Online Social Networks and Media*, Elsevier B.V., 1, 1–13.

Murthy, D., and Gross, A. J. (2017). "Social media processes in disasters: Implications of emergent technology use." *Social Science Research*, Elsevier Ltd, 63, 356–370.

Nelson, J. K., and Brewer, C. A. (2017). "Evaluating data stability in aggregation structures across spatial scales: revisiting the modifiable areal unit problem." *Cartography and Geographic Information Science*, Taylor & Francis, 44(1), 35–50.

Noy, I. (2016). "Tropical storms: The socio-economics of cyclones." *Nature Climate Change*, Nature Publishing Group, 6(4), 343–345.

Nyaupane, N., Bhandari, S., Rahaman, M. M., Wagner, K., Kalra, A., Ahmad, S., and Gupta, R. (2018). "Flood Frequency Analysis Using Generalized Extreme Value Distribution and Floodplain Mapping for Hurricane Harvey in Buffalo Bayou." *World Environmental and Water Resources Congress 2018: Watershed Management, Irrigation and Drainage, and Water Resources Planning and Management - Selected Papers from the World Environmental and Water Resources Congress 2018*.

Olteanu, A., Vieweg, S., and Castillo, C. (2015). "What to Expect When the Unexpected Happens: Social Media Communications Across Crises." *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing -*

CSCW '15, 994–1009.

Panagiotopoulos, P., Barnett, J., Bigdeli, A. Z., and Sams, S. (2016). "Social media in emergency management: Twitter as a tool for communicating risks to the public." *Technological Forecasting and Social Change*, Elsevier Inc., 111, 86–96.

Pasch, R. J., Penny, A. B., and Berg, R. (2019). "Hurricane Maria." *National Hurricane Center Tropical Cyclone Report*.

Polisciuc, E., Maçãs, C., Assunção, F., and Machado, P. (2016). "Hexagonal Gridded Maps and Information Layers: a Novel Approach for the Exploration and Analysis of Retail Data." *SIGGRAPH ASIA 2016 Symposium on Visualization on - SA '16*, ACM Press, New York, New York, USA, 1–8.

Portnov, B. A., Dubnov, J., and Barchana, M. (2007). "On ecological fallacy, assessment errors stemming from misguided variable selection, and the effect of aggregation on the outcome of epidemiological study." *Journal of Exposure Science and Environmental Epidemiology*, 17(1), 106–121.

Potter, K. M., Koch, F. H., Oswalt, C. M., and Iannone, B. V. (2016). "Data, data everywhere: detecting spatial patterns in fine-scale ecological information collected across a continent." *Landscape Ecology*, Springer Netherlands, 31(1), 67–84.

Qadir, J., Ali, A., ur Rasool, R., Zwitter, A., Sathiaseelan, A., and Crowcroft, J. (2016). "Crisis analytics: big data-driven crisis response." *Journal of International Humanitarian Action*, Journal of International Humanitarian Action, 1(1), 12.

Quarantelli, E. L. (2003). "A Half Century Of Social Science Disaster Research:Selected Major Findings And Their Applicability." *Disaster Research Center Preliminary Papers*.

Raue, S., Azzopardi, L., and Johnson, C. W. (2013). "#Trapped! Social Media Search System Requirements for Emergency Management Professionals." *Proceedings of the 36th international ACM SIGIR conference on Research and development in information retrieval - SIGIR '13*, 1073–1076.

Reed, D. A., Powell, M. D., and Westerman, J. M. (2010). "Energy Infrastructure Damage Analysis for Hurricane Rita." *Natural Hazards Review*, 11(3), 102–109.

Reibel, M., and Agrawal, A. (2007). "Areal Interpolation of Population Counts Using Pre-classified Land Cover Data." *Population Research and Policy Review*, 26(5–6), 619–633.

Reuter, C., Hughes, A. L., and Kaufhold, M. A. (2018). "Social Media in Crisis Management: An Evaluation and Analysis of Crisis Informatics Research." *International Journal of Human-Computer Interaction*, Taylor & Francis, 34(4), 280–294.

Reuter, C., and Kaufhold, M.-A. (2017). "Fifteen years of social media in emergencies: A retrospective review and future directions for crisis Informatics." *Journal of Contingencies and Crisis Management*, 41–57.

Reuter, C., and Kaufhold, M. A. (2018). "Fifteen years of social media in emergencies: A retrospective review and future directions for crisis Informatics." *Journal of Contingencies and Crisis Management*, 26(1), 41–57.

Reynard, D., and Shirgaokar, M. (2019). "Harnessing the power of machine learning: Can Twitter data be useful in guiding resource allocation decisions during a natural disaster?" *Transportation Research Part D: Transport and Environment*, Elsevier, (xxxx), 1–15.

Rogers, E. M. (2001). "The digital Divide." *Convergence*.

Roshan, M., Warren, M., and Carr, R. (2016). "Understanding the use of social media by organisations for crisis communication." *Computers in Human Behavior*, Elsevier Ltd, 63, 350–361.

Rufat, S., Tate, E., Emrich, C. T., and Antolini, F. (2019). "How Valid Are Social Vulnerability Models?" *Annals of the American Association of Geographers*, Routledge, 109(4), 1131–1153.

Saib, M.-S., Caudeville, J., Carre, F., Ganry, O., Trugeon, A., and Cicolella, A. (2014). "Spatial Relationship Quantification between Environmental, Socioeconomic and Health Data at Different Geographic Levels." *International Journal of Environmental Research and Public Health*, 11(4), 3765–3786.

Sakaki, T., Okazaki, M., and Matsuo, Y. (2010). "Earthquake Shakes Twitter Users: Real-time Event Detection by Social Sensors." *Proceedings of the 19th International Conference on World Wide Web*, 851–860.

Samuels, R., and Taylor, J. E. (2019a). “Applied Methodology for Identifying Hurricane-Induced Social Media Signal Changes in Vulnerable Populations.” *Computing in Civil Engineering 2019*, Proceedings.

Samuels, R., and Taylor, J. E. (2019b). “The Impact of Geographic Scale on Identifying Different Social Media Behavior Extremes in Crisis Research.” *2019 Winter Simulation Conference (WSC)*, IEEE, 830–841.

Samuels, R., Taylor, J., and Mohammadi, N. (2018a). “The Sound of Silence: Exploring How Decreases in Tweets Contribute to Local Crisis Identification.” *Proceedings of the 15th ISCRAM Conference*, (May), 1–9.

Samuels, R., Taylor, J., and Mohammadi, N. (2018b). “The sound of silence: Exploring how decreases in tweets contribute to local crisis identification.” *Proceedings of the International ISCRAM Conference*.

Sedgwick, P. (2014). “Spearman’s rank correlation coefficient.” *BMJ (Online)*.

Shelton, T., Poorthuis, A., Graham, M., and Zook, M. (2014). “Mapping the data shadows of Hurricane Sandy: Uncovering the sociospatial dimensions of ‘big data.’” *Geoforum*, Elsevier Ltd, 52, 167–179.

Shelton, T., Poorthuis, A., and Zook, M. (2015). “Social media and the city: Rethinking urban socio-spatial inequality using user-generated geographic information.” *Landscape and Urban Planning*, Elsevier B.V., 142, 198–211.

Spence, P. R., Lachlan, K. A., Lin, X., and del Greco, M. (2015). “Variability in Twitter Content Across the Stages of a Natural Disaster: Implications for Crisis Communication.” *Communication Quarterly*, 2015, 63(2), 171–186.

Spence, P. R., Lachlan, K. A., and Rainear, A. M. (2016). “Social media and crisis research: Data collection and directions.” *Computers in Human Behavior*, Elsevier Ltd, 54, 667–672.

Starbird, K., Maddock, J., Orand, M., Achterman, P., and Mason, R. M. (2014). “Rumors, False Flags, and Digital Vigilantes: Misinformation on Twitter after the 2013 Boston Marathon Bombing.” *iConference 2014 Proceedings*.

Stumpf, M. P. H., and Porter, M. A. (2012). “Critical Truths About Power Laws.” *Science*, 335(6069), 665–666.

Takahashi, B., Tandoc, E. C., and Carmichael, C. (2015). "Communicating on Twitter during a disaster: An analysis of tweets during Typhoon Haiyan in the Philippines." *Computers in Human Behavior*, Elsevier Ltd, 50, 392–398.

Tang, L., Zhang, Y., Dai, F., Yoon, Y., Song, Y., and Sharma, R. S. (2017). "Social Media Data Analytics for the U.S. Construction Industry: Preliminary Study on Twitter." *Journal of Management in Engineering*, 33(6), 1–15.

Tapia, A. H., and Moore, K. (2014). "Good Enough is Good Enough: Overcoming Disaster Response Organizations' Slow Social Media Data Adoption." *Computer Supported Cooperative Work: CSCW: An International Journal*, 23(4–6), 483–512.

Thatcher, J., Bergmann, L., Ricker, B., Rose-Redwood, R., O'Sullivan, D., Barnes, T. J., Barnesmoore, L. R., Beltz Imaoka, L., Burns, R., Cinnamon, J., Dalton, C. M., Davis, C., Dunn, S., Harvey, F., Jung, J. K., Kersten, E., Knigge, L. D., Lally, N., Lin, W., Mahmoudi, D., Martin, M., Payne, W., Sheikh, A., Shelton, T., Sheppard, E., Strother, C. W., Tarr, A., Wilson, M. W., and Young, J. C. (2015). "Revisiting critical GIS." *Environment and Planning A*, 48(5), 815–824.

The Federal Emergency Management Agency. (2016). "Damage assessment operations manual - A guide to assessing damage and impact." 121.

Toepke, S. L. (2018a). "Minimum Collection Period for Viable Population Estimation from Social Media." (Gistam), 138–147.

Toepke, S. L. (2018b). "Leveraging Elasticsearch and Botometer to Explore Volunteered Geographic Information." (May).

Twitter, I. (2018). "Twitter Announces Third Quarter 2018 Results." *PRNewswire*.

U.S. Census Bureau. (2016). *Decennial Census 2010*.

Villegas, C., Wowk, K., and Shelton, K. (2018). *Rethinking Disaster Recovery and Mitigation Funding in the Wake of Hurricane Harvey*. Houston, Texas.

Virkar, Y., and Clauset, A. (2014). "Power-law distributions in binned empirical data."

Annals of Applied Statistics, 8(1), 89–119.

Vuong, Q. H. (1989). "Likelihood Ratio Tests for Model Selection and Non-Nested Hypotheses." *Econometrica*.

Wagner, C. H. (1982). "Simpson's Paradox in Real Life." *The American Statistician*.

Wang, Q., and Taylor, J. E. (2015). "Process Map for Urban-Human Mobility and Civil Infrastructure Data Collection Using Geosocial Networking Platforms." *Journal of Computing in Civil Engineering*, 30(2), 04015004-1–11.

Wang, Q., and Taylor, J. E. (2016a). "Patterns and limitations of urban human mobility resilience under the influence of multiple types of natural disaster." *PLoS ONE*, 11(1), 1–14.

Wang, Q., and Taylor, J. E. (2016b). "Process Map for Urban-Human Mobility and Civil Infrastructure Data Collection Using Geosocial Networking Platforms." *Journal of Computing in Civil Engineering*, 30(2), 04015004.

Wang, Y., and Taylor, J. E. (2018). "Coupling sentiment and human mobility in natural disasters: a Twitter-based study of the 2014 South Napa Earthquake." *Natural Hazards*, Springer Netherlands, 92(2), 907–925.

Wang, Y., Wang, Q., and Taylor, J. E. (2017). "Aggregated responses of human mobility to severe winter storms: An empirical study." *PLoS ONE*, 12(12), 1–15.

Wang, Z., Lam, N. S. N., Obradovich, N., and Ye, X. (2019). "Are vulnerable communities digitally left behind in social responses to natural disasters? An evidence from Hurricane Sandy with Twitter data." *Applied Geography*, Elsevier Ltd, 108(April), 1–8.

Wehrens, R., Hovde, K., and Hiemstra, P. (2019). "Package ' pls .'"

Wurman, J., and Kosiba, K. (2018). "The role of small-scale vortices in enhancing surface winds and damage in Hurricane Harvey (2017)." *Monthly Weather Review*, 146(3), 713–722.

Xiao, Y., Huang, Q., and Wu, K. (2015). "Understanding social media data for disaster management." *Natural Hazards*, Springer Netherlands, 79(3), 1663–1679.

Xu, Z., Zhang, H., Sugumaran, V., Choo, K.-K. R., Mei, L., and Zhu, Y. (2016). "Participatory sensing-based semantic and spatial analysis of urban emergency events using mobile social media." *EURASIP Journal on Wireless Communications and Networking*, EURASIP Journal on Wireless Communications and Networking, 2016(1), 44.

Yang, K.-C., Varol, O., Davis, C. A., Ferrara, E., Flammini, A., and Menczer, F. (2019). "Arming the public with AI to counter social bots." *Human Behavior and Emerging Technologies*.

Yang, S., Chung, H., Lin, X., Lee, S., Chen, L., Wood, A., Kavanaugh, A. L., Sheetz, S. D., Shoemaker, D. J., and Fox, E. a. (2013). "PhaseVis 1 : What , When , Where , and Who in Visualizing the Four Phases of Emergency Management Through the Lens of Social Media." *Proceedings of the 10th International ISCRAM Conference*, (May), 912–917.

Yap, J. B. H., Chow, I. N., and Shavarebi, K. (2019). "Criticality of Construction Industry Problems in Developing Countries: Analyzing Malaysian Projects." *Journal of Management in Engineering*, 35(5).

Zook, M., Graham, M., Shelton, T., and Gorman, S. (2010). "Volunteered Geographic Information and Crowdsourcing Disaster Relief: A Case Study of the Haitian Earthquake." *World Medical & Health Policy*, 2(2), 6–32.

Zou, L., Lam, N. S. N. N., Shams, S., Cai, H., Meyer, M. A., Yang, S., Lee, K., Park, S.-J. J., and Reams, M. A. (2019). "Social and geographical disparities in Twitter use during Hurricane Harvey." *International Journal of Digital Earth*, Taylor & Francis, 12(11), 1300–1318.

Zou, L., Lam, N. S. N., Shams, S., Cai, H., Meyer, M. A., Yang, S., Lee, K., Park, S. J., and Reams, M. A. (2018). "Social and geographical disparities in Twitter use during Hurricane Harvey." *International Journal of Digital Earth*, Taylor & Francis, 0(0), 1–19.

Zurich, and Global Disaster Preparedness Center. (2018). "Houston and Hurricane Harvey: a call to action."